

**Estimation of the  $W$ +jets, top and QCD backgrounds in  
the search for supersymmetry with jets, missing  
transverse momentum and tau leptons**

Diploma thesis at the ATLAS experiment

by  
Felix Bühner

December 1, 2011



ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG  
FAKULTÄT FÜR MATHEMATIK UND PHYSIK



# Albert-Ludwigs-Universität Freiburg

FAKULTÄT FÜR MATHEMATIK UND PHYSIK

## **Estimation of the $W$ +jets, top and QCD backgrounds in the search for supersymmetry with jets, missing transverse momentum and tau leptons**

Diploma thesis at the ATLAS experiment

by

Felix Bühner

Date handed in: December 1, 2011

Supervised by: Prof. Dr. K. Jakobs & Prof. Dr. J. Dingfelder



## Zusammenfassung

Für die Suche nach Supersymmetrie am LHC sind Endzustände mit Tau-Leptonen durch die stärkere Kopplung an die dritte Generation für große Teile des Parameterraums von großem Interesse. Im Rahmen einer Suche nach Supersymmetrie mit Tau-Leptonen im Endzustand werden in dieser Arbeit Methoden vorgestellt um die Untergründe von  $W$ +jets, Top-Quarks und QCD-Multijet Produktion zu bestimmen. Sowohl Prozesse mit echten Tau-Leptonen und echter fehlender transversaler Energie durch nicht-wechselwirkende Teilchen, als auch Prozesse mit fehlidentifizierten Tau-Kandidaten oder fehlender transversaler Energie durch Detektoreffekte, bilden Untergründe für diese Suche. Für die vorgestellte Analyse werden Daten von  $pp$ -Kollisionen bei einer Schwerpunktsenergie von  $\sqrt{s} = 7$  TeV benutzt, die im Jahr 2011 mit dem ATLAS-Detektor aufgezeichnet wurden. Die gesamte integrierte Luminosität des analysierten Datensatzes beträgt  $\mathcal{L} = 2.05 \text{ fb}^{-1}$ .

Der Zerfall von  $W$ -Bosonen bildet den dominanten Untergrund dieser Suche. Fehlende transversale Energie kann hierbei durch nicht im Detektor nachgewiesene Neutrinos erzeugt werden. Anhand verschiedener Kontrollregionen wird dieser Untergrund aus Daten abgeschätzt.  $W$ -Bosonen können direkt produziert werden oder tauchen als Zerfallsprodukte von Top-Quarks auf. Diese beiden Beiträge werden getrennt abgeschätzt.

Es wird erwartet, dass der Untergrund von QCD-Multijet Produktion klein ist im Vergleich zu den restlichen betrachteten Untergründen. Um unabhängig von der Monte-Carlo Simulation zu sein wird dessen Größenordnung jedoch mit einer datengestützten Methode abgeschätzt.

Abschließend werden die Resultate der Suche nach supersymmetrischen Teilchen, basierend auf der in dieser Arbeit vorgestellten Abschätzung der Untergründe, zusammengefasst.

## Abstract

Final states with tau leptons hold promise for sensitivity to the search for supersymmetry at the LHC due to the enhanced couplings to the third generation for large parts of the parameter space. In the context of an inclusive search for supersymmetry with tau leptons in the final state, methods are presented to estimate the backgrounds from  $W$ +jets, top quarks and QCD-multijet production. Processes involving real tau leptons and real missing transverse energy from non-interacting particles, as well as processes with misidentified tau candidates or missing transverse energy due to detector effects are possible backgrounds for this search. In the presented analysis  $pp$ -collision data at a center-of-mass-energy of  $\sqrt{s} = 7$  TeV recorded by the ATLAS detector in 2011 are used. The analysis is based on a dataset with an integrated luminosity of  $\mathcal{L} = 2.05 \text{ fb}^{-1}$ .

The dominant background comes from  $W$  boson decays. Missing transverse energy in the event can be caused by neutrinos that leave the detector without interacting.  $W$  bosons can either be directly produced in association with jets or in the decay of top quarks. The two contributions are estimated separately.

The background from QCD-multijet production is expected to be small compared to the backgrounds from  $W$ +jets and top production. A data-driven estimate is performed to be independent of the Monte-Carlo simulation.

In conclusion the results of the search for supersymmetric particles, based on the background estimation methods presented in this thesis, are summarized.

# Contents

---

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>                                       | <b>1</b>  |
| <b>2</b> | <b>Theory</b>   | <b>3</b>  |
| 2.1      | The Standard Model of particle physics . . . . .          | 3         |
| 2.1.1    | The particle spectrum . . . . .                           | 3         |
| 2.1.2    | Mathematical description of the Standard Model . . . . .  | 5         |
| 2.1.3    | The tau lepton . . . . .                                  | 11        |
| 2.1.4    | Limitations of the Standard Model . . . . .               | 11        |
| 2.2      | Supersymmetry . . . . .                                   | 15        |
| 2.2.1    | MSSM . . . . .  | 16        |
| 2.2.2    | Supersymmetry breaking . . . . .                          | 20        |
| 2.3      | Physics of $pp$ -collisions . . . . .                     | 23        |
| 2.3.1    | SUSY at the LHC . . . . .                                 | 23        |
| <b>3</b> | <b>The ATLAS experiment at the LHC</b>                    | <b>25</b> |
| 3.1      | The Large Hadron Collider . . . . .                       | 25        |
| 3.2      | The ATLAS detector . . . . .                              | 26        |
| 3.2.1    | Inner detector components . . . . .                       | 28        |
| 3.2.2    | Calorimeters . . . . .                                    | 29        |
| 3.2.3    | Muon system . . . . .                                     | 30        |
| 3.2.4    | Triggering . . . . .                                      | 31        |
| <b>4</b> | <b>Object reconstruction and identification</b>           | <b>33</b> |
| 4.1      | Jet reconstruction . . . . .                              | 33        |
| 4.1.1    | B-tagging . . . . .                                       | 33        |
| 4.2      | Tau reconstruction and identification . . . . .           | 34        |
| 4.3      | Reconstruction of the missing transverse energy . . . . . | 35        |
| 4.4      | Reconstruction of electrons . . . . .                     | 36        |
| 4.5      | Reconstruction of muons . . . . .                         | 37        |

|          |   |           |
|----------|---|-----------|
| 4.6      | Overlap removal . . . . .   | 37        |
| 4.7      | Important variables . . . . .   | 38        |
| <b>5</b> | <b>Event selection</b>  | <b>39</b> |
| 5.1      | Data and Monte Carlo samples . . . . .                                  | 39        |
| 5.1.1    | ATLAS data sample . . . . .   | 39        |
| 5.2      | Event selection . . . . .   | 41        |
| 5.2.1    | Trigger . . . . .   | 41        |
| 5.2.2    | Data quality criteria . . . . .   | 41        |
| 5.2.3    | Channel definition . . . . .  | 42        |
| 5.2.4    | Background suppression requirements . . . . .                           | 44        |
| 5.3      | Overview of the full selection . . . . .                                | 46        |
| 5.3.1    | QCD event weighting using tau misidentification probabilities . . . . . | 48        |
| <b>6</b> | <b>Estimation of the <math>W</math>+jets and top backgrounds</b>        | <b>51</b> |
| 6.1      | $W$ +jets / top background with real tau leptons . . . . .              | 52        |
| 6.1.1    | Estimation of the QCD background in $CR_1$ . . . . .                    | 53        |
| 6.1.2    | Separation of $W$ +jets from top events . . . . .                       | 57        |
| 6.2      | $W$ +jets / top background with misidentified tau candidates . . . . .  | 72        |
| <b>7</b> | <b>Estimation of the QCD-multijet background</b>                        | <b>75</b> |
| 7.1      | Overview of the method . . . . .  | 76        |
| 7.2      | Modification of the method . . . . .                                    | 79        |
| 7.3      | Validation of the method . . . . .                                      | 84        |
| 7.3.1    | Check for correlations . . . . .  | 84        |
| 7.3.2    | Test of the robustness . . . . .  | 87        |
| 7.4      | Results . . . . .   | 88        |
| <b>8</b> | <b>Systematic uncertainties</b>   | <b>93</b> |
| 8.1      | Determination of the systematic uncertainties . . . . .                 | 93        |
| 8.1.1    | Jet energy scale and resolution . . . . .                               | 94        |
| 8.1.2    | Tau energy scale . . . . .  | 95        |
| 8.1.3    | Tau misidentification probability rate and efficiency . . . . .         | 95        |
| 8.1.4    | $b$ -tagging efficiency scale factor and uncertainty . . . . .          | 95        |
| 8.1.5    | Systematic uncertainties for light leptons . . . . .                    | 96        |
| <b>9</b> | <b>Results</b>  | <b>99</b> |
| 9.1      | Exclusion limits . . . . .  | 103       |

|                                   |            |
|-----------------------------------|------------|
| <b>10 Conclusion</b>              | <b>107</b> |
| <b>A Used Monte Carlo samples</b> | <b>109</b> |
| <b>Bibliography</b>               | <b>113</b> |
| <b>List of Figures</b>            | <b>117</b> |
| <b>List of Tables</b>             | <b>121</b> |
| <b>Acknowledgments</b>            | <b>123</b> |



# Chapter 1 Introduction

---

Everything we know today about matter and the forces between it, is summarized in a well-established framework, the Standard Model of particle physics. It describes all known elementary particles and their interactions, but, in spite of its success in the description of currently observable phenomena of particle physics, it has several shortcomings and does not hold at high energies. Therefore it can only be regarded as a low energy approximation of an underlying theory.

One possible model to address the open questions in the Standard Model is Supersymmetry, an attractive model that introduces a symmetry between fermions and bosons. In supersymmetric extensions of the Standard Model every particle has at least one superpartner differing in spin by  $1/2$  but with the same other quantum numbers. Since no supersymmetric particles have yet been observed, supersymmetry has to be broken to allow for different masses of particles and their superpartners. For many supersymmetric scenarios the coupling to the fermions of the third generation is enhanced with respect to the fermions of the first and second generation, thus final states with tau leptons hold promise for sensitivity to searches for supersymmetric particles.

The Large Hadron Collider, a particle accelerator located at CERN near Geneva, allows for a search for particles that were not accessible with former experiments, due to the very high center-of-mass energy of up to 14 TeV.

A crucial task of all searches for new physics is the understanding of the background due to already known processes. In this thesis an inclusive search for supersymmetry in final states with tau leptons, missing transverse energy and jets is presented. The main emphasis is on the estimation of the backgrounds. Background contributions from the production of  $W$  bosons in association with jets, top quarks and QCD-multijet events are considered.  $W$ +jets and top production are important backgrounds due to the tau leptons that can be produced in the decay of  $W$  bosons. Missing transverse energy in these events can be caused by neutrinos that leave the detector without interacting. In addition to events with real tau leptons, also events where a jet or light lepton is misidentified as a tau lepton contribute to the total background from  $W$ +jets and top.

The background from QCD-multijet events is expected to be small compared to the contri-

butions from  $W$ +jets and top, but it is crucial to estimate this background from data due to the high production cross-section of these processes.

The thesis is structured in the following way: First an introduction to the Standard Model of particle physics and its supersymmetric extension is given in chapter 2. Subsequently chapter 3 provides an overview of the experimental setup, followed in chapter 4 by a description of the methods for reconstructing the physics objects used in this analysis. The object selection is presented in chapter 5. The main part of this thesis is the estimation of the  $W$ +jets and top backgrounds in chapter 6 and the estimation of the background from QCD-multijet production in chapter 7. In chapter 8 the systematic uncertainties on the background estimates are discussed and chapter 9 shows the obtained results.

# Chapter 2 Theory

---

This section gives a short introduction to the Standard Model of particle physics, the motivation for searching for physics beyond the Standard Model and concludes with the basic concepts of the supersymmetric extension.

## 2.1 The Standard Model of particle physics

Everything known today about elementary particles and their interactions is described in an extremely successful framework called “the Standard Model of particle physics”. An educational introduction to the Standard Model can be found for example in [1]. In this chapter, only an overview of the concepts of the theory, as it is needed for this thesis, is given.

### 2.1.1 The particle spectrum

The particle content of the Standard Model, summarized in table 2.1, can be divided into two classes: fermions, which carry spin  $1/2$ , and bosons with integer spin.

#### Fermions

The fermions - also called matter particles - can again be divided into two classes: *quarks*, the constituents of the protons and neutrons, and *leptons*, like the well-known electron. The difference between quarks and leptons is the fact that quarks carry *color charge* and take part in strong interactions while leptons do not. Both types of fermions, quarks and leptons, are arranged in three generations with the second and third generation particles being “heavier copies” of their first-generation counterparts, differing only in their masses. The first generation in the quark-sector consists of the up- and down quark ( $u$  and  $d$ ), followed by the second generation charm and strange quarks ( $c$  and  $s$ ) and the third generation top and bottom quarks ( $t$  and  $b$ ). The first quark of each generation is called “up-type” quark and has an electric charge of  $(+2/3)^1$ , the second is a “down-type” quark with electric charge

---

<sup>1</sup>in units of the electric charge of the electron  $e = 1.6 \times 10^{-19} \text{ C}$ .

|              | spin | charge  | Generation |           |            |
|--------------|------|---------|------------|-----------|------------|
|              |      |         | I          | II        | III        |
| leptons      | 1/2  | 1       | $e$        | $\mu$     | $\tau$     |
|              |      | 0       | $\nu_e$    | $\nu_\mu$ | $\nu_\tau$ |
| quarks       | 1/2  | 2/3     | $u$        | $c$       | $t$        |
|              |      | -1/3    | $d$        | $s$       | $b$        |
| gauge bosons | 1    | 0       | $\gamma$   |           |            |
|              |      | 0       | $g$        |           |            |
|              |      | 0       | $Z^0$      |           |            |
|              |      | $\pm 1$ | $W^\pm$    |           |            |

Table 2.1: Overview of the particle spectrum of the Standard Model.

$(-1/3)$ .

In the lepton sector, the first family is comprised of the electron ( $e$ ) and its associated neutrino ( $\nu_e$ ). The two heavier copies of the electron, the muon ( $\mu$ ) and the tau lepton ( $\tau$ ), together with their corresponding neutrinos complete the lepton sector. Electron, muon and tau lepton have an electric charge of  $-1$  by definition. The neutrinos are electrically neutral. For all particles there also exists an anti particle with identical mass and spin, but opposite values for all other non-zero quantum numbers.

The visible matter is exclusively formed by fermions from the first generation.

### Bosons

The interactions of the particles are mediated via the exchange of gauge bosons. Every fundamental force described by the Standard Model has its associated boson(s).

- The photon  $\gamma$  is the massless mediator of the electromagnetic interaction and couples to the electric charge of the particles, thus only the quarks and, charged leptons and the  $W^\pm$  “feel” the electromagnetic force.
- Eight massless gluons  $g$  carry the strong force, which binds the quarks in the nucleons and ultimately in the atomic nucleus. Gluons only couple to particles carrying color charge, the quarks and gluon themselves. Leptons and other bosons do not take part in the strong interactions

| interaction     | couples to      | relative strength   | mediator(s)         | mass(es) (GeV)   |
|-----------------|-----------------|---------------------|---------------------|--|
| strong          | color charge    | 1                   | 8 gluons ( $g$ )    | 0  |
| electromagnetic | electric charge | $1 \times 10^{-3}$  | photon ( $\gamma$ ) | 0  |
| weak            | weak charge     | $1 \times 10^{-14}$ | $W^\pm, Z$          | $m_{W^\pm} = 80.399 \pm 0.023$<br>$m_Z = 91.1876 \pm 0.0021$ |

Table 2.2: Interactions in the Standard Model, their relative strengths and the corresponding mediators. The relative strengths depend on the scale and the distance and thus should not be taken literally. Adapted from [2].

- The three vector bosons  $W^\pm$  and  $Z$  mediate the weak interaction that all particles are subjected to.

The relative strengths of the forces, together with the associated gauge bosons are summarized in table 2.2.

**The Higgs boson** is the only particle of the Standard Model that has not been experimentally observed. It is needed to explain the masses of the massive gauge bosons  $W^\pm$  and  $Z$  and all other Standard Model particles through Yukawa coupling to the Higgs field.

### 2.1.2 Mathematical description of the Standard Model

The three forces of the Standard Model are described by quantum field theories (QFT). The oldest of these quantum field theories is the description of the electromagnetic interaction between electrically charged particles, called quantum electrodynamics (QED). Quantum chromodynamics (QCD) describes the strong interaction of particles with color charge and the weak force is described in the Glashow-Salam-Weinberg model by unification of the electromagnetic and the weak interaction.

#### Quantum electrodynamics (QED)

Quantum electrodynamics describes the interaction of (electrically) charged particles through the exchange of photons. Since it is the easiest description of a QFT in the Standard Model, it is used here to illustrate the general concept.

In classical mechanics the motion of a particle in space can be described by introducing a Lagrangian  $L$  that is a function of the (generalized) coordinates  $q_i$  and their time derivatives

$\hat{q}_i$ . In quantum field theory, the subject of interest is not the particle itself, but the fields  $\phi_i(x, y, z, t)$  and the particles are regarded as excited states of these fields.

In analogy to classical mechanics, quantum field theory starts with a Lagrangian density  $\mathcal{L}$ , which is a function of the fields  $\phi_i$  and their spatial  $(x, y, z)$  and time  $(t)$  derivatives. The Lagrangian density of a particle with spin  $1/2$  and mass  $m$ , to describe fermions, is

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi. \quad (2.1)$$

The  $\partial_\mu$  are the time and space derivatives and  $\gamma^\mu$  denote the Dirac matrices, which can be found for example in [1]. This Lagrangian density is invariant under *global phase transformations* of the form

$$\psi \rightarrow e^{i\theta}\psi, \quad \theta \in \mathbb{R}. \quad (2.2)$$

If a phase factor, different at different space-time points ( $\theta \rightarrow \theta(x)$ ) is applied to the Lagrangian density, this is called a *local phase transformation*. It shows that  $\mathcal{L}$  is not invariant under these transformations.

Demanding  $\mathcal{L}$  to also be invariant under local phase transformations leads to a changed Lagrangian density  $\mathcal{L}'$  that introduces a new field  $A_\mu$ , transforming according to the rule  $A_\mu \rightarrow A_\mu + \partial_\mu\left(-\frac{\theta}{q}\right)$ . It includes a term describing a coupling of this new field  $A_\mu$  to  $\psi$ :

$$\mathcal{L}' = \mathcal{L} - (q\bar{\psi}\gamma^\mu\psi)A_\mu. \quad (2.3)$$

For this changed Lagrangian density to be locally gauge invariant one more term, describing the *free* motion of  $A_\mu$ , has to be added:

$$\mathcal{L}'' = \mathcal{L}' - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (2.4)$$

This Lagrangian density is now locally gauge invariant if the field  $A_\mu$  is massless and hence can be identified as the field of the photon  $\gamma$ , coupling to  $\psi$  via the electric charge  $q$ .  $F_{\mu\nu}$  is the *field strength tensor* known from the classical description of electrodynamics (see for example [3]).

The phase transformation defined in equation (2.2) can be thought of as the multiplication of  $\psi$  with a  $1 \times 1$  unitary matrix  $U$

$$\psi \rightarrow U\psi \quad \text{with } U^\dagger U = 1. \quad (2.5)$$

The group of all such matrices is  $U(1)$ , hence the underlying symmetry is called  *$U(1)$  gauge invariance*.

This general scheme of starting with a Lagrangian density for the particles that are to be described and then demanding local gauge invariance is common to all quantum field theories

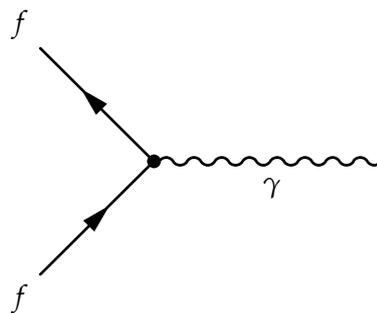


Figure 2.1: The elementary vertex of QED.

in the Standard Model.

The new Lagrangian density  $\mathcal{L}''$  can be regarded as comprised of two terms, the free Lagrangian and the interaction terms  $\mathcal{L}_{int}$ . To use this Lagrangian to describe particles (in this case fermions) and their interactions, it is translated into a set of rules, the so-called *Feynman Rules*. These rules can then be used to calculate the transition probability of a given initial state to a final state with the help of perturbation theory. The Feynman Rules are not only a very powerful tool for calculating the transition probabilities, but can also be used as a graphical representation (The Feynman Diagrams) of the process under study.<sup>2</sup> While Feynman Diagrams are a good pictorial representation of a physics process, they should not be regarded as showing the true motion of the particles in space-time.

Figure 2.1 shows the elementary interaction of QED that is obtained from  $\mathcal{L}''$ . With the help of this single vertex all processes between fermions through the exchange of photons can be described. To calculate the full transition probability for a given physics process, for example electron-muon scattering  $e\mu \rightarrow e\mu$ , an infinite number of these diagrams need to be calculated and added up. Fortunately, in many cases higher order diagrams only lead to corrections to the leading order ("tree-level") calculation. The leading order and an example for a next-to-leading order correction for electron-muon scattering is shown in figure 2.2.

### Quantum chromodynamics (QCD)

Quantum chromodynamics describes the strong interaction of particles carrying color charge. Each quark flavor comes in three different colors - red, blue and green. The free Lagrangian density is the same as in equation 2.1, again to describe particles with spin 1/2, but in QCD  $\psi$  now stands for the three-component column vector

---

<sup>2</sup>The time axis for all Feynman diagrams in this document is horizontal.

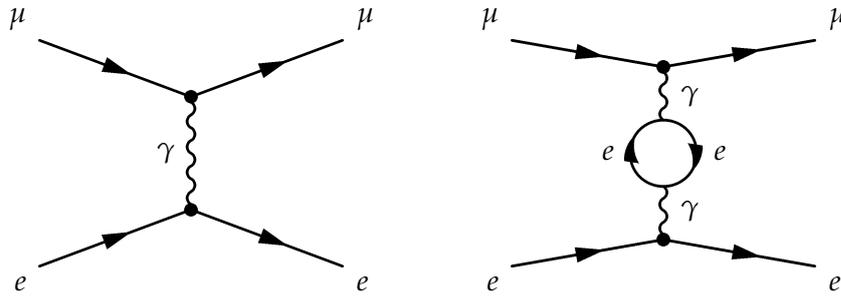


Figure 2.2: Feynman diagrams for electron-muon scattering. Tree level (left), example for a next to leading order correction (right).

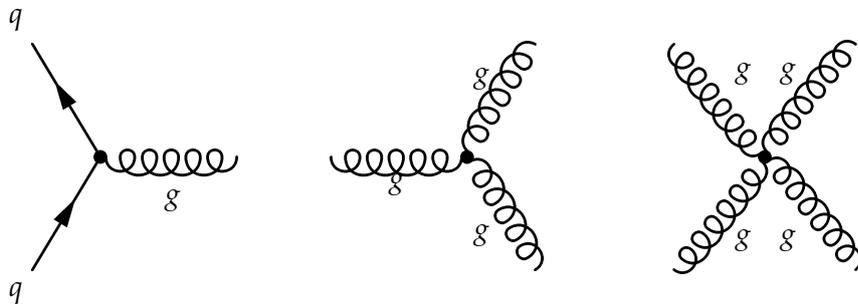


Figure 2.3: Possible vertices of QCD.

$$\psi = \begin{pmatrix} \psi_r \\ \psi_b \\ \psi_g \end{pmatrix}. \quad (2.6)$$

The Lagrangian density is again invariant under global transformations of the form  $\psi \rightarrow U\psi$ , this time  $U$  is a unitary  $3 \times 3$  matrix, hence the underlying symmetry of QCD is  $SU(3)$ .

In analogy to QED, local gauge invariance is demanded leading to a modified Lagrangian density that includes eight gauge fields representing the eight gluons in QCD. The fact that gluons themselves carry color charge leads to three different possible vertices in QCD. In addition to the  $qqg$  vertex, the QCD-equivalent of figure 2.1, a three-gluon vertex and a four-gluon vertex is possible. The three vertices are shown in figure 2.3.

The three-gluon and four-gluon vertices lead to an increase of the coupling strength if

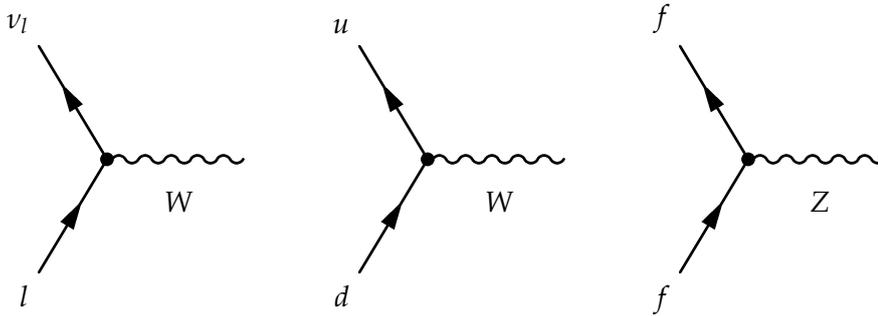


Figure 2.4: Vertices of the weak interaction.

the distance between two quarks is increased. The result of this is that quarks and gluons<sup>3</sup> cannot be observed as free particles, but are bound in color neutral hadrons. This is called *confinement*.

### Electroweak unification

The weak interaction, most prominently seen in the  $\beta$ -decay of atomic nuclei, can only be explained together with the electromagnetic force in a combined *electroweak* theory. This has first been accomplished by Glashow, Salam and Weinberg by defining the *hypercharge*  $Y$ .  $Y$  links the electric charge  $Q$  with the *isospin operators*  $T^i$ , the generators of the weak force, by:

$$Q = T^3 + \frac{1}{2}Y. \quad (2.7)$$

The hypercharge generates a  $U(1)$  symmetry as described in section 2.1.2, while the three isospin-operators have a  $SU(2)$ <sup>4</sup> structure.

As a consequence this yields three gauge fields  $W_\mu^i$  from the  $SU(2)$ -part and one gauge field  $B_\mu$  from the  $U(1)$  symmetry. Mixing of the  $W^{1,2}$  fields leads to two charged, massive bosons  $W^\pm = \frac{1}{2}(W_\mu^1 \mp iW_\mu^2)$ . The neutral bosons  $W_\mu^3$  and  $B_\mu$  mix to one massive field  $Z_\mu$  and one massless field  $A_\mu$ , the photon field.

$$A_\mu = B_\mu \cos \theta_w + W_\mu^3 \sin \theta_w \quad (2.8)$$

$$Z_\mu = -B_\mu \sin \theta_w + W_\mu^3 \cos \theta_w \quad (2.9)$$

with the electroweak mixing angle  $\theta_w$ .

The three vertices of the weak interaction are shown in figure 2.4.

<sup>3</sup>The partons

<sup>4</sup>The group of unitary 2x2 matrices  $U$  with  $\det(U) = 1$

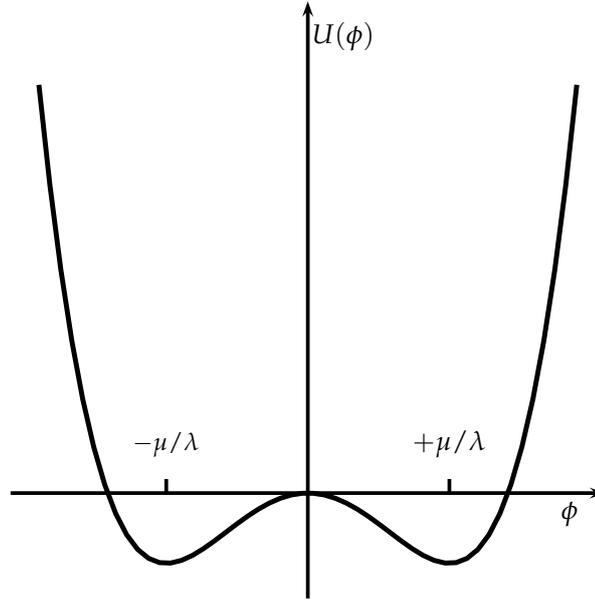


Figure 2.5: Graph of  $U(\phi)$  (Equation 2.10).

### The Higgs mechanism

The Lagrangian density describing the Standard Model forbids terms that involve masses of the gauge bosons. To generate the masses of the massive gauge bosons  $W^\pm$  and  $Z$  a concept called *spontaneous symmetry breaking* is applied.

The general idea is to introduce two scalar fields  $\phi_i$  to the Lagrangian that generate a potential term of the form

$$U(\phi) = -\frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda^2\phi^4 \quad (2.10)$$

in a way that the Lagrangian itself is invariant under rotations in the  $\phi_1, \phi_2$  space. Figure 2.5 shows the form of this potential for one of the  $\phi_i$ . Demanding the ground state of the field to be stable, it shows that while the Lagrangian is invariant under this symmetry (this can be thought of as a reflection at the vertical axis of figure 2.5), the ground state (the minimum of the potential) of the system is not. This is called *spontaneous symmetry breaking*.

By breaking the  $SU(2) \times U(1)$  symmetry of the electroweak description in the Standard Model, four massless *Goldstone bosons* appear in the theory. They can be regarded as four degrees of freedom of which three can be used to generate masses for the three massive gauge bosons  $W^\pm$  and  $Z^0$ . The remaining degree of freedom leads to one massive, scalar, electrically neutral boson, the Higgs boson.

The Higgs boson is also responsible for the masses of the fermions, which are assumed

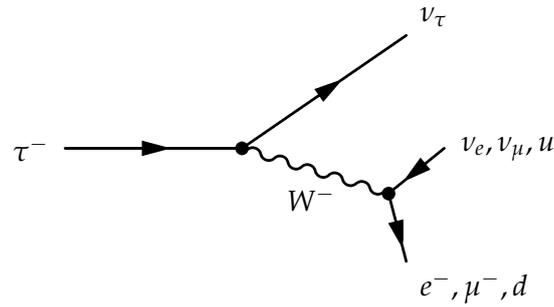


Figure 2.6: Feynman diagram for the decay of a tau lepton into light leptons or quarks.

to have Yukawa couplings to the Higgs field. These couplings generate mass terms in the Lagrangian density after the spontaneous symmetry breaking.

### 2.1.3 The tau lepton

Considering that the aim of this thesis is to present a search involving tau leptons, the most important properties of the tau lepton are discussed - especially the differences to the lighter charged leptons, the electron and muon, concerning the experimental observations.

With a mass of  $m_\tau = (1776.84 \pm 0.17)$  MeV [4], roughly 2000 times the mass of the electron, the tau lepton is by far the heaviest lepton, exceeding the mass of the lightest mesons. It has a mean lifetime of  $(2.1970190 \pm 0.0000021) \times 10^{-6}$  s. In contrast to the muon, which can only decay via  $\mu \rightarrow \nu_\mu + e + \nu_e$  into an electron and two neutrinos, the tau lepton can also decay into quarks which then hadronize forming a spray of particles. This is shown in figure 2.6. The branching ratio into hadrons is about 65% and the different hadronic decay modes of the tau lepton can be classified according to the number of charged daughter particles. Table 2.3 shows the different decay modes of the tau lepton with one charged track (1-prong) and three charged tracks (3-prongs) and their branching fractions. Since the decay products of non-hadronic tau decays cannot be distinguished from directly produced light leptons (*prompt* leptons), the term “tau lepton” in the context of this analysis always implies a hadronic decay.

### 2.1.4 Limitations of the Standard Model

Although the Standard Model is a very successful theory that withstood all experimental tests over a large range of energies, there are limitations in the Standard Model. In the following these limitations are discussed.

| decay channel                      | No. of prongs | BR [%] |
|------------------------------------|---------------|--------|
| $\pi^- \nu_\tau$                   | 1             | 16.8   |
| $\pi^- \pi^0 \nu_\tau$             | 1             | 39.4   |
| $\pi^- \pi^0 \pi^0 \nu_\tau$       | 1             | 14.4   |
| $\pi^- \pi^+ \pi^- \nu_\tau$       | 3             | 14.4   |
| $\pi^- \pi^+ \pi^- \pi^0 \nu_\tau$ | 3             | 7.1    |

Table 2.3: Most prominent hadronic decay-channels of the tau lepton and their branching fractions.

### Gravitation

In addition to the three forces described by the Standard Model, there is also the gravitational force, whose strength is about 40 order of magnitudes smaller than the electromagnetic force, but still has a large influence at large mass scales, due to its infinite range. Within the Standard Model there is no explanation of the gravitational force between particles. Since gravitation is so small, it only starts to play a role at energies orders of magnitudes larger than what can be accomplished with particle physics experiments.

### The hierarchy problem

Although the Higgs boson is yet to be discovered, its mass is known to be between 114 GeV and 155 GeV at the 95% confidence level. The lower bound is set by direct searches done at the LEP<sup>5</sup> experiments at CERN, the upper bound is set only in 2011 by the ATLAS collaboration [5]. Indirect electroweak measurements exclude Higgs masses of 185 GeV and above. Since the Higgs boson is a scalar particle, its mass can include large corrections by loop diagrams and hence can be thought of as a *bare mass*,  $m_{H,\text{bare}}$ , and the possible loop corrections  $\Delta m_H^2$ :

$$m_H^2 = m_{H,\text{bare}}^2 + \Delta m_H^2. \quad (2.11)$$

All fermions and the electroweak gauge bosons can appear in loop diagrams and contribute to the mass of the Higgs boson. The contributions from fermions are proportional to the coupling to the Higgs field and the mass of the fermion.

These loop corrections are quadratically divergent. This can be avoided by introducing a

---

<sup>5</sup>Large Electron Positron

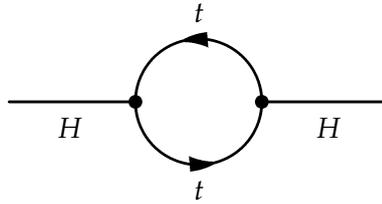


Figure 2.7: Contribution to the Higgs mass from a loop diagram involving the top quark.

cut-off scale  $\Lambda$  that sets the scale up to which the theory should remain valid:

$$\Delta m_{H, \text{fermion}}^2 \propto \lambda_f \cdot \left[ -\Lambda^2 + m_f^2 \ln \left( \frac{\Lambda}{m_f} \right) \right] + \dots, \quad (2.12)$$

where  $\lambda_f$  and  $m_f$  denote the coupling of the fermion to the Higgs field and its mass.

Figure 2.7 depicts the contribution from the top quark to the Higgs mass, which is the largest fermionic contribution due to the high mass of the top quark.

Loop corrections from bosons contribute with a relative minus sign compared to the contributions from fermions, but are considerably smaller and cancel the effect only to a small fraction.

The relatively small Higgs mass can be explained by “fine-tuning” the value of  $m_{H, \text{bare}}$  to  $\Delta m_f$ . If the Standard Model is valid up to the Planck scale  $M_P \sim 10^{19}$  GeV, the scale at which the effect of gravitation starts to play a role, the value of the bare Higgs mass has to be fine-tuned to 34 orders of magnitude, which seems to be very unnatural.

### Unification of the couplings

The idea of a grand unified theory<sup>6</sup> is, that at a certain energy scale all interactions of the Standard Model can be unified to a single interaction with a larger symmetry which is broken at smaller energies, leading to the interactions observed and explained in the Standard Model.

This also means that the *coupling constants*, describing the strengths of each of the interactions, which change depending on the energy scale, meet at the GUT scale. Within the Standard Model the running of the coupling can be calculated and it is seen that the coupling constants do not meet. This is shown in figure 2.8. The calculation of the evolution of the coupling constants with the energy can be found for example in [6].

<sup>6</sup>GUT

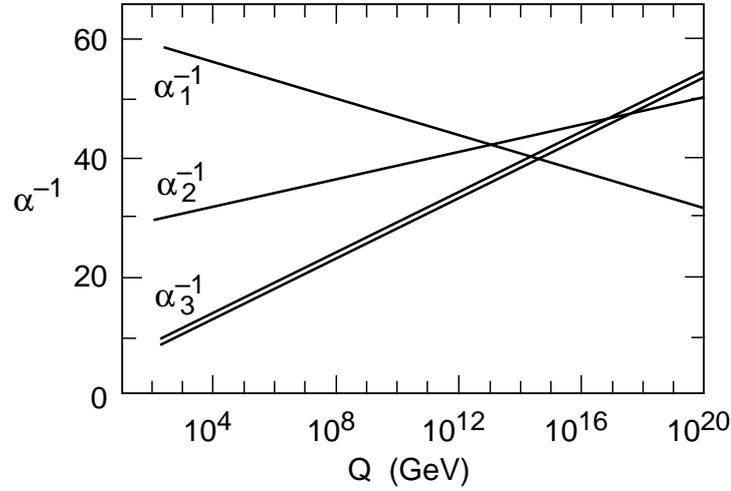


Figure 2.8: Evolution of the  $SU(3) \times SU(2) \times U(1)$  gauge couplings to high energy scales, using the one-loop renormalization group equations of the Standard Model. The double line for  $\alpha_3$  indicates the current experimental error in this quantity; the errors in  $\alpha_1$  and  $\alpha_2$  are too small to be visible. Taken from [6].

### Dark matter

Observations of the rotational speed of galaxies show that there has to be a type of massive matter in our universe that does not interact electromagnetically, hence *dark matter*. It makes up about 85% of all matter in the universe. The only particles in the Standard Model that could explain the dark matter are the neutrinos, which are known to have small, but non-zero masses. By observation of the dark matter, however, it has been found that dark matter develops large-scale structures that neutrinos could not form due to the high speed they are traveling at. This means that there is no candidate in the Standard Model that can explain dark matter.

In addition, cosmological observations of the expansion of the universe have shown that  $\sim 74\%$  of the energy density of the universe is caused by something entirely different called *dark energy*.

Figure 2.9 shows the contributions to the energy density of the universe split up in contributions from dark energy, dark matter and baryonic matter. The Standard Model only describes the distinguished part.

The non-unification of the couplings at large energy and the hierarchy problem are more of an esthetical than a real problem of the theory, since the hierarchy problem can be overcome by fine-tuning as described and the idea of the unification of the coupling strengths at high energies silently assumes that a grand unified theory exists.

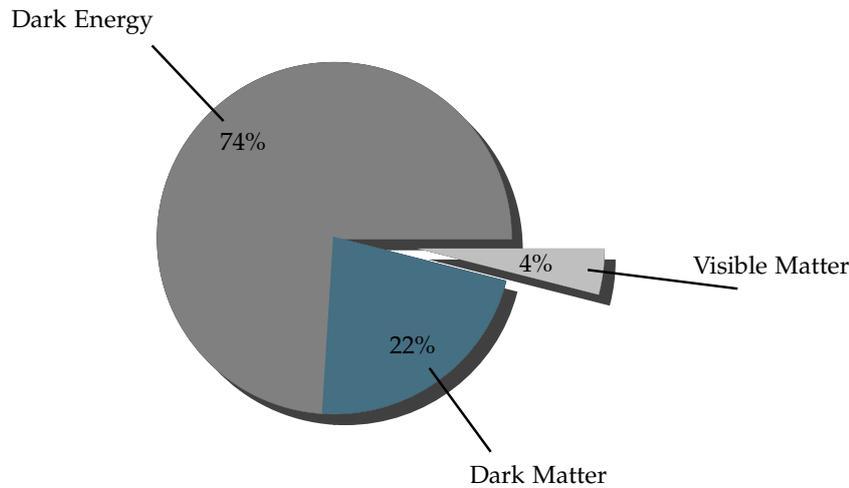


Figure 2.9: Representation of the energy density of the universe, split up in contributions from dark energy, dark matter and visible baryonic matter. Only the later is described by the Standard Model.

The clearest hint that the Standard Model cannot be an inclusive theory of particle physics is the observation of dark matter. A theory, no matter how successful, that is only able to describe a small fraction of the energy content in the universe can only be part of the answer.

## 2.2 Supersymmetry

The success of the Standard Model in describing all experiments can not be ignored when looking for physics beyond the Standard Model. A theory called *supersymmetry* (SUSY) can be used to remedy several of the limitations of the Standard Model. It is an extension of the Standard Model with an additional symmetry between bosons and fermions. Building up on the Standard Model gives the possibility to incorporate phenomena studied in the context of the Standard Model into the theory. A good introduction to the principles of supersymmetry can be found in [7] and [8]. Again, only a brief introduction to the topic is given.

The main idea behind supersymmetry is the introduction of an operator  $Q$  that, when applied to a fermionic state, yields a bosonic state and vice versa:

$$Q |\text{boson}\rangle \propto |\text{fermion}\rangle , \quad (2.13)$$

$$Q |\text{fermion}\rangle \propto |\text{boson}\rangle . \quad (2.14)$$

Particles related by equations 2.13 and 2.14 are called *superpartners*.

The operator  $Q$  is the generator of the SUSY transformation and is, in contrast to the scalar generators of the symmetries described in the Standard Model, a two-component spinor and can therefore link particle states with different spins.

To incorporate such a theory in a four-dimensional quantum field theory,  $Q$  and its hermitian conjugate  $Q^\dagger$  must satisfy the following commutation and anticommutation relations:

$$\{Q, Q^\dagger\} = P^\mu, \quad (2.15)$$

$$\{Q^\dagger, Q^\dagger\} = \{Q, Q\} = 0 \text{ and} \quad (2.16)$$

$$[P^\mu, Q^\dagger] = [P^\mu, Q] = 0 \quad (2.17)$$

with the four-momentum operator  $P^\mu$ .

From these relations it follows that superpartners must have the same eigenvalues of the squared mass operator  $-P^2$  and therefore equal masses and also have to be members of the same gauge group leading to the same quantum numbers.

### 2.2.1 MSSM

The **Minimal Supersymmetric Standard Model** (MSSM) is the implementation of this idea with the smallest number of particles needed. Every particle in the Standard Model (with an exception in the Higgs sector discussed below) has exactly one superpartner differing by spin of  $1/2$ . The particles and their superpartners (or rather the fields and corresponding superfields) are arranged in *supermultiplets*. In the MSSM two types of supermultiplets are distinguished:

- Chiral supermultiplets, comprising of a spin-1/2 field and a complex scalar field,
- Gauge supermultiplets, built from a spin-1 field and a spin-1/2 field.

The chiral supermultiplets in the MSSM are shown in table 2.4, table 2.5 summarizes the gauge supermultiplets.

The superpartners of the fermions are called *sfermions*<sup>7</sup>, denoted  $\tilde{f}$ , and have spin 0<sup>8</sup>. The superpartners of the bosons are named by adding *-ino* at the end of the name<sup>9</sup>. They have

<sup>7</sup>for *scalar* fermions

<sup>8</sup>the superpartner of the electron  $e$  for example is the selectron  $\tilde{e}$

<sup>9</sup>the superpartner of the  $W$  boson is called Wino

| Names             | spin 0                      | spin 1/2                        |
|-------------------|-----------------------------|---------------------------------|
| squarks, quarks   | $(\tilde{u}_L \tilde{d}_L)$ | $(u_L d_L)$                     |
|                   | $\tilde{u}_R^*$             | $u_R^+$                         |
|                   | $\tilde{d}_R^*$             | $d_R^+$                         |
| sleptons, leptons | $(\tilde{\nu} \tilde{e}_L)$ | $(\nu_L, e_L)$                  |
|                   | $\tilde{e}_R$               | $e_R^+$                         |
| Higgs, higgsinos  | $(H_u^+ H_u^0)$             | $(\tilde{H}_u^+ \tilde{H}_u^0)$ |
|                   | $(H_d^0 H_d^-)$             | $(\tilde{H}_d^0 \tilde{H}_d^-)$ |

Table 2.4: Chiral supermultiplets in the Minimal Supersymmetric Standard Model. Adapted from [7].

| Names           | spin 1/2                     | spin 1       |
|-----------------|------------------------------|--------------|
| gluino, gluon   | $\tilde{g}$                  | $g$          |
| winos, W bosons | $\tilde{W}^\pm, \tilde{W}^3$ | $W^\pm, W^3$ |
| bino, B boson   | $\tilde{B}$                  | $B$          |

Table 2.5: Gauge supermultiplets in the Minimal Supersymmetric Standard Model. Adapted from [7].

spin 1/2.

In addition to that not one but two scalar Higgs-fields are needed to give mass to up-type and down-type particles and to avoid triangular anomalies. This leads to eight degrees of freedom. Three of them are absorbed in the masses of the gauge bosons  $W^\pm$  and  $Z$ . The remaining degrees of freedom results in five physical Higgs bosons:

- 2 scalar:  $h, H^0$ ,
- 1 pseudoscalar:  $A$ ,
- 2 charged:  $H^\pm$ .

Analogous to the mixing of the  $W^3$  and the  $B$  in the electroweak unification (see section 2.1.2), their spin-1/2 superpartners, the winos and bino, mix. In contrast to the elec-

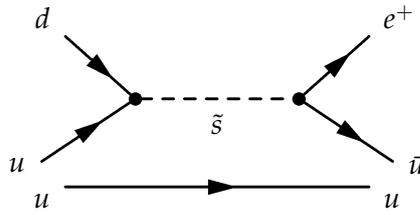


Figure 2.10: Possible Feynman Diagram of a proton-decay if both lepton and baryon number-violating terms are allowed.

troweak mixing, additional particles, the Higgsinos, also have spin-1/2 and hence the observed particles are mixed from winos, bino and the higgsinos. These masseigenstates<sup>10</sup> can be split up in two different classes:

- the electrically charged *charginos*  $\chi_{1,2}^{\pm}$
- the electrically neutral *neutralinos*  $\chi_{1,2}^0$ .

### R-parity conservation

There is, however, a problem with the theory, namely the inclusion of vertices which violate the conservation of lepton and baryon number, which would allow the proton to decay into  $e^+ + \pi^0$  shown in figure 2.10. While there is no theoretical limitation, the lifetime of the proton has been measured to be larger than  $10 \times 10^{31}$  years.

To avoid this caveat a new multiplicative quantum number,  $R_P$  (*R-parity* defined as

$$R_P = (-1)^{3(B-L)+2s} \quad (2.18)$$

for each particle with baryon number  $B$ , lepton number  $L$  and spin  $s$ . By demanding this quantum number to be conserved, processes that violate both the conservation of lepton and baryon number are forbidden. The  $R$ -parity is defined in the way that each Standard Model particle has  $R_P = +1$ , while all superpartners have  $R_P = -1$ . This has two major implications when considering SUSY models with  $R$ -parity conservation:

- Superpartners are always produced in pairs
- The lightest supersymmetric particle (LSP) is stable, since it cannot decay into lighter Standard Model particles.

<sup>10</sup>the gauginos

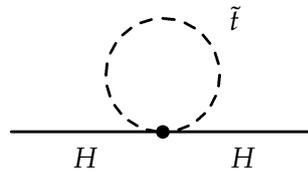


Figure 2.11: Contribution to the Higgs mass from a loop involving the superpartner of the top quark, the *stop*  $\tilde{t}$ .

### Solutions to the limitations of the Standard Model

Supersymmetry has solutions for some of the limitations of the Standard Model described in section 2.1.4.

- When going from global SUSY invariance to local SUSY invariance, as done in some models, an additional gauge field is needed. This can be interpreted as a spin 3/2 gravitino that forms a supermultiplet with its superpartner, the spin 2 graviton. Gravitation can therefore be explained as a mere consequence on demanding local SUSY invariance.
- Cancellation of loop corrections to the bare Higgs mass. In addition to the fermion loops in figure 2.7, also the superpartners contribute via loop diagrams to the Higgs mass. Due to the difference of the spins of the superpartners the corrections cancel each other. Figure 2.11 shows the loop Feynman diagram for the superpartner of the top quark. This cancellation happens in every order of the calculation.
- Unification of the coupling constants. In supersymmetric models, the three coupling constants can meet at high energies. This is shown in figure 2.12. The running of the coupling constants is sensitive to the different parameters of the SUSY-model and the possibility of the unification is not a global characteristic of supersymmetry itself.
- A dark matter candidate. In supersymmetric scenarios with  $R$ -parity conservation, the lightest supersymmetric particle is stable. If the LSP would take part in the strong or electromagnetic interaction it would have become bound to nuclei and atoms. This would have been found in searches for anomalous isotopes. Hence the lightest supersymmetric particle must be only weakly interacting.

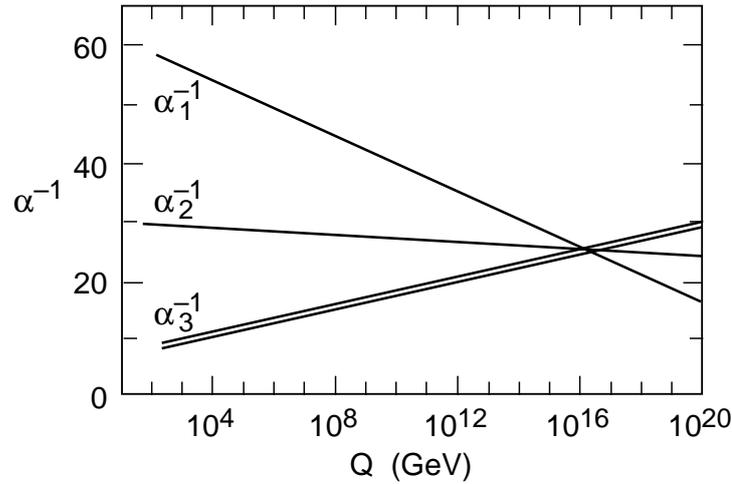


Figure 2.12: Evolution of the  $SU(3) \times SU(2) \times U(1)$  gauge couplings to high energy scales, using the one-loop renormalization group equations of the supersymmetric generalization of the Standard Model. From [6].

### The role of tau leptons in SUSY searches

In the MSSM sfermions of the same family can mix, which gives rise to contributions leading to a reduction of the sfermion mass eigenvalue. The size of the contribution is proportional to the mass of the fermion and the ratio of the vacuum expectation values for the two Higgs doublets,  $\tan\beta$ . In the first two generations the mixing gives only a small contribution due to the small masses. The masses of the left and right handed sfermions are almost degenerated. In the third generation, the mixing of the left and right handed states lead to two mass eigenstates that can have a large difference in mass. The lighter of the two mass eigenstates is in general lighter than the sfermions from the first and second generation. Especially in scenarios with large  $\tan\beta$ , sfermions of the third generation are produced more frequently than the superpartners of the lighter fermions, leading to more tau leptons in the final states compared to lighter leptons. For large  $\tan\beta$ , the coupling of neutralinos and charginos to tau leptons is enhanced and the decay channels to light leptons might even be closed.

### 2.2.2 Supersymmetry breaking

Among the implications of the commutator relations in equations 2.15–2.17 is the demand that superpartners have equal mass. The fact that no supersymmetric particle has yet been found at an experiment, although for example the superpartner of the electron  $\tilde{e}$ , a very light electrically charged particle with spin 1, should be quite easy to detect, leads to the conclusion that supersymmetry cannot be a perfect symmetry, but has to be broken. This

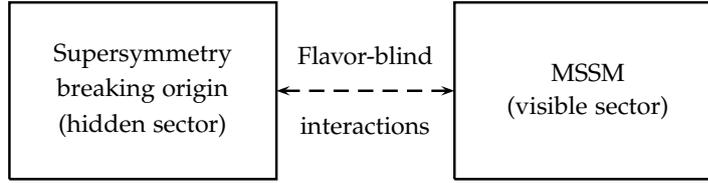


Figure 2.13: The presumed schematic structure for supersymmetry breaking. Adapted from [7].

implies that the cancellation in the loop corrections to the Higgs mass are not perfect, since they depend on the masses of the particles in the loop. For supersymmetry to offer a valid solution to the hierarchy problem, the masses of the supersymmetric partners should not be too large. This is called “soft”-SUSY breaking and the Lagrangian can be written as

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{SOFT}} , \quad (2.19)$$

where  $\mathcal{L}_{\text{SUSY}}$  is the unbroken part, invariant under SUSY transformations, and  $\mathcal{L}_{\text{SOFT}}$  contains all possible terms that can be added but still protect the Higgs doublets from quadratic mass renormalization.

This of course raises the question on how this breaking occurs and the amount of free parameters in SUSY increases considerably due to the now unknown masses of the supersymmetric particles.

Different models of SUSY-breaking have been studied and are discussed for example in [8]. They all have in common that the origin of the supersymmetry breaking happens in a *hidden sector* and the breaking is mediated by flavor-blind interactions between this hidden sector and the MSSM. This is shown schematically in figure 2.13. The different models of SUSY-breaking use different particles to mediate between the hidden and the visible sector. A short introduction to two of these models, the gauge-mediated supersymmetry breaking (GMSB) and minimal supergravity (mSugra) is given below.

### Gauge-Mediated SUSY Breaking (GMSB)

As the name gauge-mediated SUSY breaking already indicates, the effects of the SUSY breaking from the hidden sector is communicated via Standard Model gauge bosons. On tree-level the supersymmetry is unbroken in this model, the breaking only occurs through loop-diagrams involving a new chiral supermultiplet called “messengers”. The model can be characterized by six parameters:

- The SUSY breaking scale  $\Lambda$ , which determines the mass scale of the SUSY particles ,

- the mass-scale of the messenger-particles  $M_{\text{mess}} > \Lambda$ ,
- the number of introduced messenger fields  $N_5$ ,
- the ratio of the vacuum expectation values of the two Higgs-doublets  $\tan \beta$ ,
- the sign of the higgsino mixing-parameter  $\text{sgn}(\mu)$ ,
- $C_{\text{grav}} \geq 1$ , the ratio of the gravitino mass to its value at the breaking scale  $\Lambda$ . This determines the lifetime of the second lightest supersymmetric particle (The NLSP).

In this analysis GMSB models in the  $(\Lambda, \tan \beta)$  plane are studied with the other parameters of the model fixed to

- $M_{\text{mess}} = 250 \text{ TeV}$ ,
- $N_5 = 3$ ,
- $\mu > 0$ ,
- $C_{\text{grav}} = 1$ .

The lightest supersymmetric particle is, independent of all parameters, a very light gravitino. The NLSP can be either the  $\chi_1^0$ , the  $\tilde{\tau}$ , or  $\tilde{e}$  and  $\tilde{\mu}$ , which leads to final states containing photons, tau leptons, electrons or muons. For large  $\tan \beta$  the NLSP is the  $\tilde{\tau}$  for most of the parameter space.

### Minimal Supergravity (mSUGRA)

Minimal supergravity is a breaking model considering local supersymmetry. The breaking occurs through gravitational interaction between the hidden sector and the MSSM. The LSP in this model is the lightest neutralino  $\chi_1^0$ . The number of free parameters is reduced by constraints on the masses at the GUT-scale. The parameters of the model are:

- The unified scalar mass at the GUT-scale  $m_0$ ,
- the unified gaugino mass at the GUT-scale  $m_{1/2}$ ,
- the unified trilinear coupling at the GUT-scale  $A_0$ ,
- the ratio of the vacuum expectation values of the two Higgs-doublets  $\tan \beta$ ,
- the sign of the higgsino mixing-parameter  $\text{sgn}(\mu)$ .

An example of a decay of a squark  $\tilde{q}$  with two tau leptons in the final state is shown in figure 2.14.

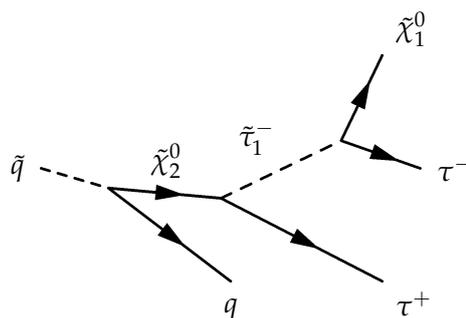


Figure 2.14: Example for a decay of a  $\tilde{q}$  involving two tau leptons.

## 2.3 Physics of $pp$ -collisions

At the Large Hadron Collider (LHC) two beams of protons are collided. The proton is a composite particle made up of quarks and gluons. Three different types of constituents (or partons) of the proton are distinguished:

- The *valence quarks*  $uud$
- gluons that are exchanged between the valence quarks and
- *sea quarks* that can be produced in pairs by gluons.

The particle content of a proton is described by a so-called *parton density function (PDF)* that specifies the probability for finding a parton  $q_i$  with the momentum fraction  $x$  of the proton, dependent on the momentum transfer  $Q^2$ .

The cross-section  $\sigma$  of a specific physics process can be calculated by factorization into the parton-parton interaction and a soft term described by the PDFs.

The true momentum fraction of the partons is unknown.

### 2.3.1 SUSY at the LHC

At hadron colliders like the LHC, the production of the superpartners of the partons, the squarks and gluinos, has a much larger cross-section compared to the production of sleptons, neutralinos and charginos. Figure 2.15 shows a Feynman diagram for the process  $pp \rightarrow \tilde{q}\tilde{q}$ . Both produced squarks decay via a cascade like the one shown in figure 2.14.

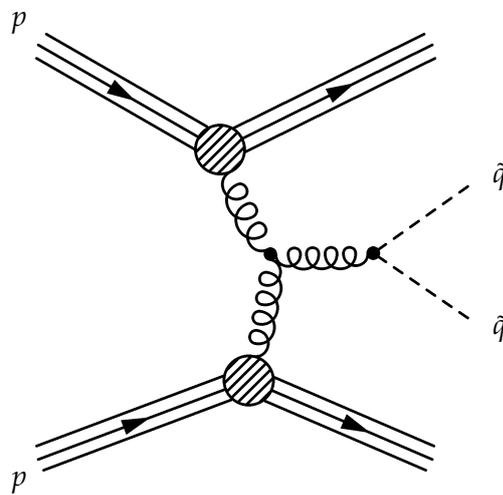


Figure 2.15: Feynman Diagram of the production of two squarks in a  $pp$ -collision.

# Chapter 3 The ATLAS experiment at the LHC

---

## 3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [9] is a particle collider located at CERN<sup>1</sup> near Geneva. It accelerates and collides protons at a center-of-mass energy of currently  $\sqrt{s} = 7 \text{ TeV}$ . Data-taking at  $\sqrt{s} = 7 \text{ TeV}$  started in 2010 and will continue until the technical stop at the end of 2012. After that an increase of the center-of-mass energy to 13 - 14 TeV is planned. It is installed in the 26.7 km long tunnel of its predecessor at CERN, the Large Electron Positron Collider (LEP).

The rate of a given physics process is not only dependent on the energy of the collision, but also on the luminosity  $L$  of the machine and can be calculated as:

$$\dot{N}_{\text{event}} = L \cdot \sigma, \quad (3.1)$$

where  $\sigma$  is the cross-section of the process under study. The design luminosity of the LHC is  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . This high luminosity can be achieved by colliding bunches of  $10^{11}$  protons every 25 ns, leading to 23 collisions per bunch crossing on average.

The pre-accelerated protons are injected into the LHC beam pipe and accelerated to their maximum energy using a high-frequency system of superconducting magnets.

In addition to colliding protons the LHC can also be operated with two beams of lead ions.

Six different experiments are built around the collision points:

- Two multi-purpose detectors, ATLAS [10] and CMS [11].
- One experiment dedicated to the measurement of  $B$ -meson decays, LHCb [12].
- ALICE [13], an experiment investigating the quark-gluon plasma in Pb-Pb collisions.
- Two smaller experiments, LHCf [14] and TOTEM [15].

---

<sup>1</sup>Conseil Européen pour la Recherche Nucléaire

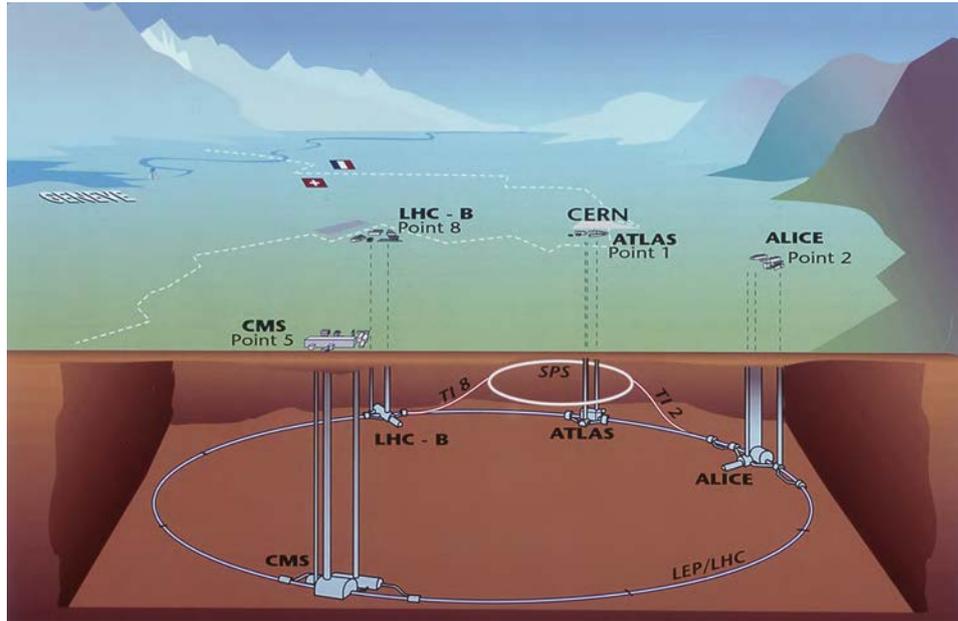


Figure 3.1: Schematic view of the Large Hadron Collider with the four large experiments, ATLAS, CMS, LHCb and ALICE [16].

Figure 3.1 shows a schematic view of the LHC, together with one of the pre-accelerators (SPS) and the four large experiments.

### 3.2 The ATLAS detector

The ATLAS<sup>2</sup> detector, shown in figure 3.2, is a cylindrical, close to  $4\pi$  detector. To measure the momenta of particles produced in  $pp$ -collisions, a hybrid magnet system consisting of:

- A solenoid magnet aligned around the beam axis, providing a 2 T axial magnetic field for the inner detector,
- a barrel and two end-cap solenoid magnets, which produce a toroidal magnetic field of approximately 1 T and 0.5 T in the central and end-cap regions respectively.

The magnet system is needed to bend the trajectories of electrically charged particles to measure their momenta.

At the ATLAS experiment a right-handed coordinate system is used to describe the position of the measured particles. The  $z$ -axis of this coordinate system is defined by the beam pipe, while the  $(x - y)$ -plane is perpendicular to the  $z$ -axis, with the  $x$ -axis pointing to the center

<sup>2</sup>A Toroidal LHC Apparatus

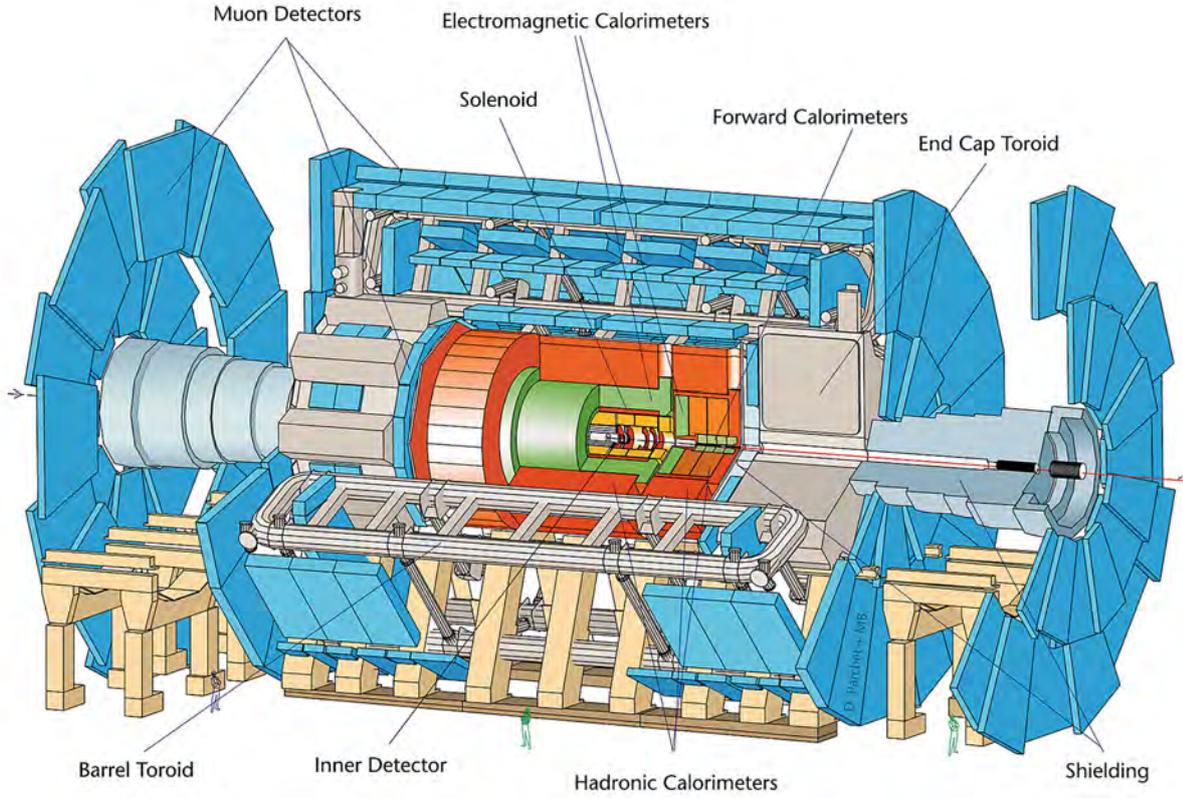


Figure 3.2: Schematic view of the full ATLAS detector [17].

of the ring and the  $y$ -axis upwards.

The *transverse momentum*,  $p_T$ , of a particle and azimuthal angles  $\phi$  are measured in the transverse ( $x - y$ )-plane. Polar angles  $\theta$  are measured with respect to the beam (or  $z$ -) axis. Instead of the polar angle  $\theta$ , the *pseudo-rapidity*,  $\eta$  can be defined as:

$$\eta = -\ln \tan \left( \frac{\theta}{2} \right), \quad (3.2)$$

which is Lorentz-invariant for massless particles.

The spatial separation  $\Delta R$ , between two objects in the detector can be calculated as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \quad (3.3)$$

In the following the different parts of the ATLAS detector components, from the inside to the outside, are described. A complete description on all detector components can be found in [10] or [18].

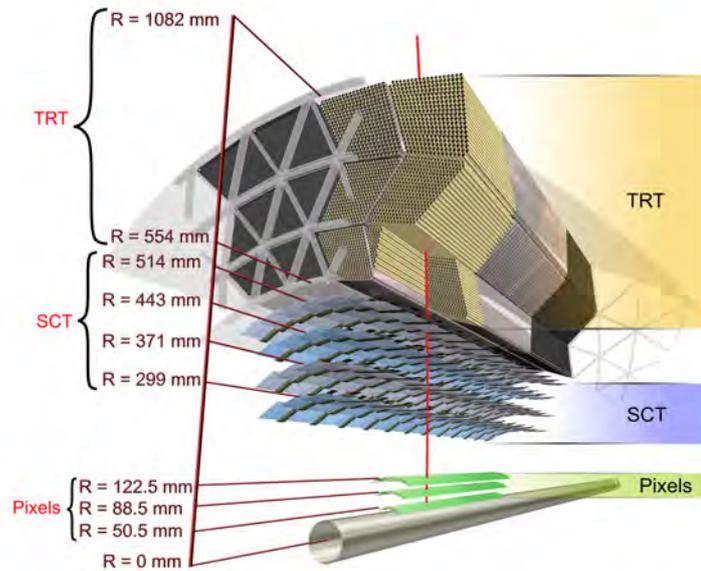


Figure 3.3: The ATLAS inner detector components. The innermost Pixel Detector followed by the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). The dimensions of the different detector components measured from the beam axis are given [19].

### 3.2.1 Inner detector components

The inner detector surrounding the beampipe is used for precise tracking of the produced charged particles. It consists of three different subdetectors, the innermost Pixel Detector, being followed by the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). The design and dimensions of the inner detector are shown in figure 3.3.

It is crucial for many physics analyses to have a good resolution for determining the interaction point. The Pixel Detector [20] consists of three layers of silicon pixels with a minimum pixel size of  $50\ \mu\text{m} \times 400\ \mu\text{m}$  for the layer closest to the interaction point. This leads to a resolution of up to  $10\ \mu\text{m}$  in  $x - y$  and  $115\ \mu\text{m}$  in  $z$ .

After the Pixel Detector a particle traverses the SCT, consisting of several layers of narrow silicon strips. By tilting the layers with respect to each other a precise measurement of the position of the traversing particle can be obtained, with resolutions of up to  $17\ \mu\text{m}$  in  $x - y$  and  $580\ \mu\text{m}$  in  $z$ .

The last subsystem of the inner detector is the Transition Radiation Tracker [21]. It consists

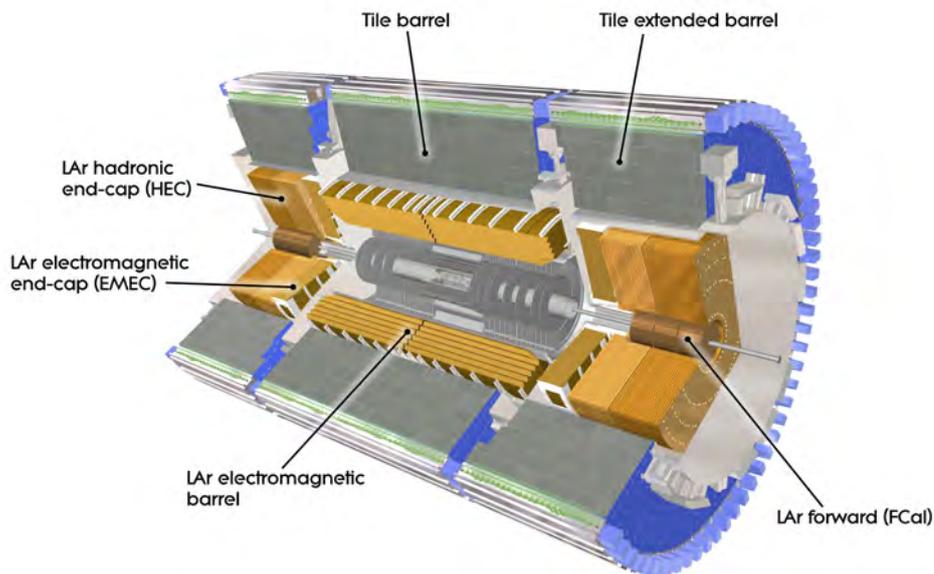


Figure 3.4: Image of the ATLAS calorimeter system [23].

of more than 298,000 straw tubes with a diameter of  $\sim 4$  mm, filled with *radiator* material that causes traversing particles to emit photons in the flight direction of the particle. These photons are then absorbed and the signal can be used for particle identification. More details can be found in [22]. The TRT has an intrinsic resolution of  $130 \mu\text{m}$  in  $x - y$ , but provides no position information along the beam-axis.

### 3.2.2 Calorimeters

In the ATLAS detector, two different types of calorimeters are used. The *electromagnetic calorimeter* measures the energy depositions of electrons and photons through electromagnetic interactions. Hadrons loose energy by deposits in the *hadronic calorimeter* mostly via strong interactions. The two types of calorimeters are shown in figure 3.4.

An electron or photon in the electromagnetic calorimeter produces a *shower* of electron-positron pairs and bremsstrahlung photons, subsequently decreasing the energy of the original particle until it is stopped. As the active detector material liquid argon is used and the layers of liquid argon are separated by read-out electrodes and absorber material (steel). The energy depositions of the shower are then summed up to obtain the energy of the particle that is to be measured. The resolution of the energy measurement in the electromagnetic calorimeter is given by

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\% , \quad (3.4)$$

where  $E$  is the energy of the particle in GeV.

Of course hadrons also lose energy when traversing the electromagnetic calorimeter, but their showers are less compact than electromagnetic showers and extend beyond the electromagnetic calorimeter into the hadronic calorimeter.

Hadronic calorimeters use strong interactions of the hadrons and the atomic nuclei of an absorber material, leading to a shower of secondary hadrons. The hadronic calorimeter at ATLAS is a system of two different sampling calorimeters, having the absorber separated from the detector, thus only sampling a fraction of the energy in the shower. The tile barrel calorimeter uses steel absorbers and scintillating tiles to measure the deposited energy. In the forward and end-cap calorimeters, a combination of copper and liquid argon is used. The resolution of the energy depositions is given by

$$\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \quad \text{for } |\eta| < 3 \text{ and} \quad (3.5)$$

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 5\% \quad \text{for } 3 < |\eta| < 5 , \quad (3.6)$$

where  $E$  is again the energy of the particle in GeV.

### 3.2.3 Muon system

The muons, due to their high mass compared to electrons, lose less energy in the calorimeters and reach the outermost layer of the detector, the muon spectrometer. As a result of the toroidal magnetic field outside of the calorimeters, the trajectories of muons are bent, allowing the measurement of the transverse momentum of the muons.

The muon spectrometer uses drift tubes and multiwire proportional chambers to detect muons traversing the three layers of the spectrometer. The three measured points can then be used to reconstruct the muon trajectory and hence its transverse momentum in the detector. To enhance the resolution of the transverse momentum of the muons, the tracks in the muon spectrometer are combined with the information from the inner detector components.

Figure 3.5 shows a summary of the signals of different types of particles in the various components of the ATLAS detector.

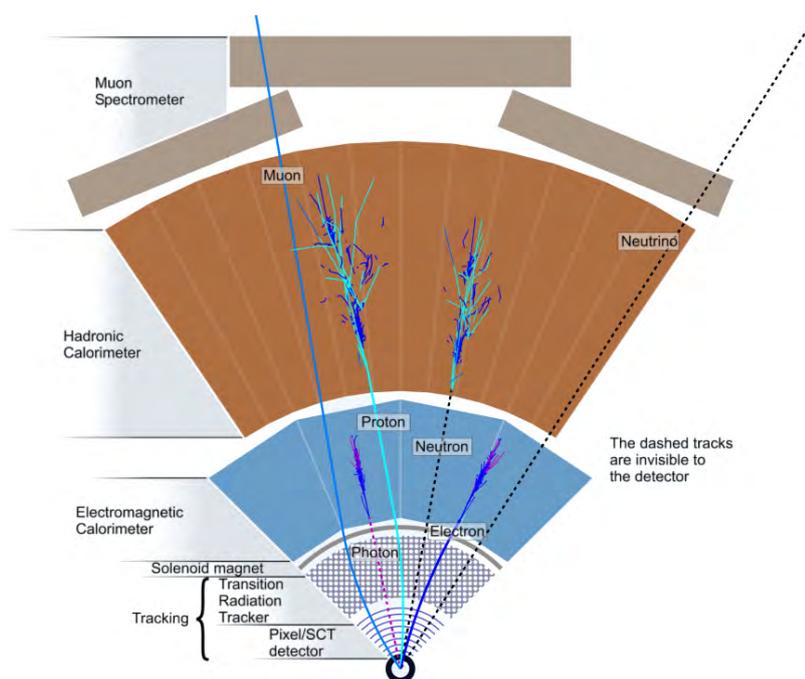


Figure 3.5: Schematic cross-section of the ATLAS detector showing the interaction of different types of particles in the various detector components [24].

### 3.2.4 Triggering

The high luminosity discussed in section 3.1 poses a problem for the data acquisition, because the interaction rate is several orders of magnitude higher than the maximal rate of data that can be recorded.

A solution to this is to select and record only events that involve processes of interest. This is called *triggering*. At ATLAS, three stages of triggers are distinct that subsequently reduce the amount of data passed to the next trigger level.

The first stage (level one) is hardware-based and uses very simple objects in the detector to be able to take very fast decisions and defines *regions of interest*.

The level two trigger partially reconstructs the event around the regions of interest.

The last step in the trigger path is the software-based event filter (EF), which uses fully reconstructed event data.

For every analysis the efficiency of the chosen trigger needs to be studied and a potential bias introduced by the trigger needs to be corrected for.



# Chapter 4 Object reconstruction and identification

---

The final state of the analysis presented in this thesis consists of at least one hadronically decaying tau lepton, missing transverse energy due to the LSP leaving the detector undetected and several high-energetic jets. Events with at least one electron or muon are vetoed.

In this chapter the reconstruction and identification of these objects, with an emphasis on jets, tau leptons and missing transverse energy, are discussed.

At the end of the chapter important kinematic variables used in the analysis are defined.

## 4.1 Jet reconstruction

As described in section 2.1.2, quarks and gluons cannot be observed as free particles because of the nature of the strong force. They hadronize and can therefore only be reconstructed as sprays of hadrons in the detector, so-called *jets*. A schematic sketch of this is shown in figure 4.1. A good introduction to the problem of defining jets from observed signals in the detector can be found in [25].

The ATLAS software uses an algorithm called anti- $k_T$  to reconstruct jets in an event, which does a successive pairing of calorimeter clusters in an iterative procedure starting with the most energetic cluster in the event. A more detailed description of the anti- $k_T$  algorithm is given in [18].

In this analysis, only reconstructed jets with  $p_T > 20 \text{ GeV}$  and  $|\eta| < 2.8$  are considered.

### 4.1.1 B-tagging

Many interesting physics processes involve  $b$ -quarks, since they are the heaviest quarks that form hadrons in the Standard Model. The heavier top quarks almost exclusively decay into a  $W$  boson and a  $b$ -quark before they can form hadrons. The identification of jets originating from a  $b$ -quark (“ $b$ -tagging”) is a technique to separate the jets from  $b$ -quarks from other jets from lighter quarks or gluons.

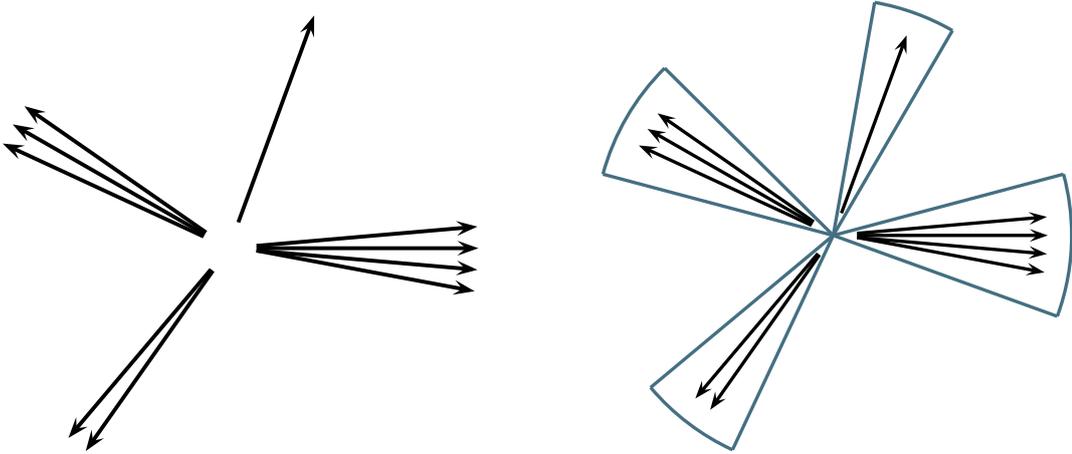


Figure 4.1: Sketch of the definition of jets (right) from the detector signals in an event (left).

This is possible due to the relatively long lifetime of  $b$ -hadrons which is in the order of  $10^{-12}$  s, depending on the specific hadron. The lifetime of the  $b$ -hadron translates to an average flight length of  $\sim 3$  mm for a  $b$ -quark with a momentum of 50 GeV, before the  $b$ -hadron decays. This flight length can be measured by reconstructing the secondary vertex from the decay of the  $b$ -hadron.

In this analysis the  $b$ -tagging algorithm JetFitterCombNN [26, 27] is used. It uses a neural network discriminator to separate  $b$ -jets from light jets. The working point used has an efficiency of 60% and a purity of 94.6% [28].

The efficiency of the tagger has been checked on data and scale factors accounting for the differences between data and Monte Carlo were published by the flavor-tagging working group at ATLAS. These scale factors are applied in the analysis.

## 4.2 Tau reconstruction and identification

The reconstruction, energy calibration and identification of hadronically decaying tau leptons is described in [29]. Only the basic idea is given here as it is needed for the presented analysis.

The signatures of hadronically decaying tau leptons are narrow, low-multiplicity jets compared to jets from prompt quarks or gluons. Throughout this analysis the following notation was chosen:

- *tau lepton* denotes the true particle or an object truth-matched to a tau lepton,
- *tau candidate* denotes an object reconstructed as a tau lepton,

- *misidentified tau candidate* denotes an object, reconstructed as but not truth-matched to a tau lepton (both light leptons or jets originating from quarks and gluons).

The seed for reconstructed tau candidates are calorimeter jets found by the anti- $k_T$  algorithm discussed above.

To discriminate between jets from prompt quarks or gluons and hadronic tau lepton decays, three different discriminators are currently implemented in the ATLAS software, based on rectangular cuts, a likelihood method and a boosted decision tree.

In this analysis the boosted decision tree classifier ( $BDT_{\tau}$ ) is used. It exploits several shower shape and tracking variables of the tau jets. Compared with jets from prompt quarks or gluons, tau jets are narrower, have less associated particles and charged tracks.  $BDT_{\tau}$  uses eleven variables to do the separation and defines three flags “*loose*”, “*medium*” and “*tight*” with a tau identification efficiency of 60%, 50% and 30%, respectively, and background rejection of 90%, 97% and 99.5%. In this analysis the *tight* flag for 1-prong and the *medium* flag for 3-prong candidates is used, where “prong” denotes the number of charged particles of the tau decay.

In addition to the identification via  $BDT_{\tau}$ , several requirements are applied to the tau candidate to obtain even better suppression of misidentified tau candidates:

- The tau candidate has to fail a dedicated electron veto. It suppresses misidentified tau leptons from electrons by a factor of 33, while having a tau identification efficiency of 75%.
- The number of tracks must be either 1 or 3 (see section 2.1.3).
- The measured sum of the charges of the tracks associated to the tau candidate,  $q$ , is required to be -1 or 1.

Tau candidates fulfilling the requirements  $p_T > 15\text{ GeV}$  and  $|\eta| < 2.5$  are used in this analysis.

### 4.3 Reconstruction of the missing transverse energy

In  $pp$ -collisions, the sum of the momenta in the  $z$ -direction is unknown because of the unknown initial momenta of the partons involved in the hard interaction (see section 2.3). The sum of all momenta in the transverse plane is expected to be zero due to momentum conservation.

The missing transverse energy  $E_T^{\text{miss}}$ , is defined as the momentum imbalance in this plane and can be caused by unseen particles, such as neutrinos or the lightest supersymmetric particle. It can be calculated by vectorially summing up the momenta of all particles detected in an event.

Details on the calculation of  $E_T^{\text{miss}}$  and its performance can be found in [30]. Only the basic concepts are discussed here.

Two components contribute to  $E_T^{\text{miss}}$ . The sum of all deposition of energy in the calorimeters  $E_{x(y)}^{\text{calo}}$  and the contributions from reconstructed muons  $E_{x(y)}^\mu$ . The transverse missing energy is calculated as

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}, \quad (4.1)$$

where

$$E_{x(y)}^{\text{miss}} = -E_{x(y)}^{\text{calo}} - E_{x(y)}^\mu. \quad (4.2)$$

The calorimeter cells used to obtain  $E_{x(y)}^{\text{calo}}$  are assigned to reconstructed high- $p_T$  physics objects to ensure the right calibration scheme. The calorimeter term of the missing energy in x and y is then obtained by summation over the separate missing energy contributions of the physics objects. The  $E_T^{\text{miss}}$  definition used in this analysis does not include a separate term for hadronically decaying tau leptons. They are treated as jets in the calculation of  $E_T^{\text{miss}}$  which is important for the discussion of the systematic uncertainties in chapter 8 in this analysis.

## 4.4 Reconstruction of electrons

To reconstruct and identify electrons at ATLAS, the information on the inner detector tracks is combined with the shower information from the electromagnetic calorimeter.

The reconstruction uses a cluster of cells in the second layer of the electromagnetic calorimeter as a seed. The electron leaves a track when traversing the inner detector. This track is associated with the seed-cluster in the calorimeter. Details on this algorithm can be found in [18]. It has an reconstruction efficiency for isolated electrons of about 93%.

Additional identification criteria are then applied to the reconstructed electron. Depending on the needed identification efficiency and background rejection, shower shape information from the electromagnetic calorimeter, the leakage into the hadronic calorimeter, the quality of the track and information from the Transition Radiation Tracker (TRT) can be used.

Identified electrons with  $p_T > 20 \text{ GeV}$  and  $|\eta| < 2.47$  are used in this analysis.

## 4.5 Reconstruction of muons

Different algorithms for the reconstruction and identification of muons at ATLAS exist [18]. The one used in this analysis combines information from the muon spectrometer and matches them to tracks from the inner detector. To quantify the quality of the match, a combined  $\chi^2$ -fit of the muon track to the measured positions in the inner detector is done with a minimum requirement on the goodness-of-fit.

Only muons with  $p_T > 20 \text{ GeV}$  and  $|\eta| < 2.4$  are considered in the selection.

## 4.6 Overlap removal

After the reconstruction and identification of all physics objects in an event, it can happen that the same object is identified more than once. Every identified tau lepton for example is also reconstructed as a jet. To avoid double-counting of objects, an overlap removal is done in the following order:

1. if an electron is within a cone of radius  $\Delta R < 0.2$  of a tau candidate, the electron is kept and the tau candidate is rejected;
2. if a muon is within a cone of radius  $\Delta R < 0.2$  of a tau candidate, the muon is kept and the tau candidate is rejected;
3. if an electron is within a cone of radius  $\Delta R < 0.2$  of a jet, the electron is kept and the jet is rejected;
4. if a tau candidate is within a cone of radius  $\Delta R < 0.2$  of a jet, the tau candidate is kept and the jet is rejected;
5. if both a jet and an electron are within  $0.2 < \Delta R < 0.4$ , the electron is assumed to be from a secondary decay within the jet and is rejected;
6. if both a jet and a muon are within  $0 < \Delta R < 0.4$ , the muon is assumed to be from a secondary decay within the jet and is rejected.

The selection requirements in the analysis, discussed in chapter 5, are then applied using the remaining objects in the event.

## 4.7 Important variables

**The transverse momentum**,  $p_T$ , of a particle is the projection of the momentum vector of a particle on the plane transverse to the beam axis.

**The effective mass**,  $m_{\text{eff}}$ , is defined as the scalar sum of the transverse components of the selected objects in an event. Since this analysis requires a tau lepton, missing transverse energy and multiple jets the effective mass is calculated as:

$$m_{\text{eff}} = p_T^\tau + E_T^{\text{miss}} + \sum p_T^{\text{jet}}, \quad (4.3)$$

where the sum runs over the leading two jets in the event. The decay of heavy particles leads to higher effective masses, hence this quantity can be used to estimate the mass scale of the decaying particles [31].

**The transverse mass**  $m_T(\tau, E_T^{\text{miss}})$ , is the invariant mass of the selected tau lepton and the missing transverse energy in the transverse plane:

$$m_T(\tau, E_T^{\text{miss}}) = \sqrt{m_\tau^2 + 2p_T^\tau \cdot E_T^{\text{miss}} \cdot (1 - \cos \Delta\phi(\tau, E_T^{\text{miss}}))}. \quad (4.4)$$

If the  $E_T^{\text{miss}}$  and the tau lepton originate from the visible decay of a massive particle (for example from a  $W \rightarrow \tau\nu$  decay), the transverse mass has an upper bound at the mass of this particle.

**The transverse thrust**  $T_T$  is an event shape variable that measures how *dijet-like* an event is. It is defined as

$$T_T = \max_{|\vec{n}_T|} \frac{\sum \vec{p}_T \cdot \vec{n}_T}{\sum p_T}. \quad (4.5)$$

where  $\vec{n}_T$  denotes the *thrust-axis*, for which  $T_T$  becomes maximal. The thrust of a perfectly balanced di-jet event is 1 along the thrust axis given by the two jets.

# Chapter 5 Event selection

---

This section documents the inclusive one-tau analysis, starting with an overview of the used data and Monte Carlo samples followed by the event selection and a discussion of the relevant backgrounds.

Since the main part of this thesis is the estimation of  $W$ +jets, top and QCD-multijet backgrounds, only an outline of the selection is given here. A more detailed documentation of the whole analysis, carried out in collaboration with a larger group, can be found in [32].

## 5.1 Data and Monte Carlo samples

To compare theoretical predictions of the Standard Model backgrounds and hypothetical signal contributions with data, one needs to rely on simulations which include both information about the simulated physics process as well as the response of the detectors.

The simulated samples for the backgrounds and the SUSY benchmark points used in this analysis are provided by the ATLAS SUSY working group. A short summary of the used background and signal samples with the associated cross-sections and the generator used is given in table 5.1 and 5.2, respectively. A full overview of all used background samples can be found in appendix A.

### 5.1.1 ATLAS data sample

For the presented analysis  $pp$  collision data collected at  $\sqrt{s} = 7$  TeV in 2011 with the ATLAS detector is used. The total recorded luminosity of this dataset is  $2.56 \text{ fb}^{-1}$ , with a total integrated luminosity available for the analysis of  $2.05 \text{ fb}^{-1}$ . The relative uncertainty on this luminosity is estimated to be 3.7%. Further information on the determination of the luminosity and the associated uncertainty can be found in [33].

| Sample name              | Generator    | $\epsilon_f \cdot \sigma$ [pb] |
|--------------------------|--------------|--------------------------------|
| $W \rightarrow e\nu$     | Alpgen/Jimmy | 10486.19                       |
| $W \rightarrow \mu\nu$   | Alpgen/Jimmy | 10483.41                       |
| $W \rightarrow \tau\nu$  | Alpgen/Jimmy | 10481.01                       |
| $t\bar{t}$ semileptonic  | MC@NLO/Jimmy | 89.36                          |
| $t\bar{t}$ full had.     | MC@NLO/Jimmy | 75.21                          |
| single $t$               | MC@NLO/Jimmy | 37.34                          |
| $Z \rightarrow ee$       | Alpgen/Jimmy | 1072.62                        |
| $Z \rightarrow \mu\mu$   | Alpgen/Jimmy | 1072.33                        |
| $Z \rightarrow \tau\tau$ | Alpgen/Jimmy | 1072.79                        |
| $Z \rightarrow \nu\nu$   | Alpgen/Jimmy | 968.15                         |
| QCD                      | Pythia       | $\sim 1.0582 \times 10^{10}$   |

Table 5.1: Monte Carlo samples used in the analysis, the generators that were used to produce the samples and the filter efficiency times cross-section for leading order (QCD), next-to-leading order ( $t\bar{t}$  and single top) or next-to-next-to-leading order ( $W$ +jets and  $Z$ +jets). More details can be found in appendix A.

| Sample name                              | Generator | $\sigma$ [pb] |
|--|-----------|---------------|
| GMSB ( $\Lambda = 30, \tan \beta = 20$ ) | Herwig++  | 1.95          |
| GMSB ( $\Lambda = 40, \tan \beta = 30$ ) | Herwig++  | 0.41          |

Table 5.2: Monte Carlo signal samples used in the analysis, the generator that were used to produce the sample and the cross-section.

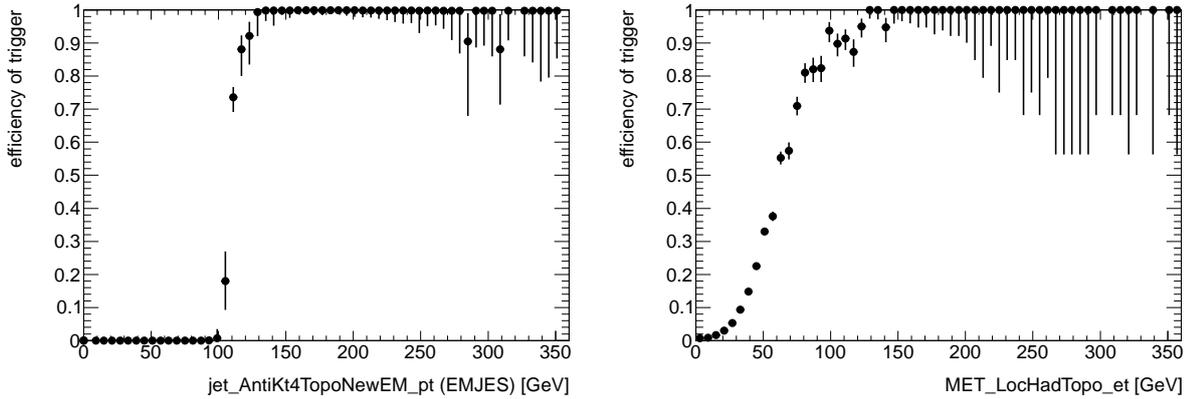


Figure 5.1: Trigger efficiency measured in data for the jet+ $E_T^{\text{miss}}$ -trigger used in the analysis versus the  $p_T$  of the leading jet (left) and  $E_T^{\text{miss}}$  (right). It can be seen, that the trigger is fully efficient after offline requirements on  $p_T^{\text{jet}1} > 130$  GeV and  $E_T^{\text{miss}} > 130$  GeV. Taken from [34].

## 5.2 Event selection

### 5.2.1 Trigger

In this analysis two triggers based on one high- $p_T$  jet and large  $E_T^{\text{miss}}$  are used. The same triggers were also used by other  $E_T^{\text{miss}}$ -based analyses in the SUSY group. In [34] a detailed study on this combination of triggers can be found in the context of the search for SUSY with jets,  $E_T^{\text{miss}}$  and no lepton.

It is concluded that the triggers are fully efficient after offline requirements on  $p_T^{\text{jet}1} > 130$  GeV and  $E_T^{\text{miss}} > 130$  GeV respectively. The trigger turn-on is shown for the  $p_T$  of the leading jet and  $E_T^{\text{miss}}$  in figure 5.1.

The Monte Carlo samples used in this analysis are known not to reproduce the trigger behavior in data well, hence only the off-line selection criteria on  $p_T^{\text{jet}1}$  and  $E_T^{\text{miss}}$  are applied to the Monte Carlo samples. Given the fact that the triggers are shown to be in the plateau-region for the used off-line selection thresholds this does not introduce a bias.

### 5.2.2 Data quality criteria

Before applying any analysis criteria, it needs to be checked that all subdetectors relevant for the analysis are operating normally and that the magnets are at full field strength. This is done by applying a so-called good run list<sup>1</sup>, whose criteria are common to all supersymmetry searches with  $E_T^{\text{miss}}$  in ATLAS.

<sup>1</sup>Run denotes a data-taking period of typically 5 - 10 hours [35].

Additional quality criteria are applied to the analysis:

**Primary vertex requirement** The analysis requires a reconstructed primary vertex in the event with at least four tracks associated with it. The reason for this requirement is to reduce the non-collision background.

**Event cleaning** The analysis vetoes events with so-called “bad” jets which are typically caused by noise or calorimeter malfunction, non-collision data or cosmic events. “Bad” jets are defined as jet objects with  $p_T > 20$  GeV that pass certain criteria commonly defined to all supersymmetry +  $E_T^{\text{miss}}$  searches at ATLAS [36].

**LAr hole treatment** During the 2011 data run, used for this analysis, a malfunction in the readout of the liquid argon calorimeter occurred. Due to the persisting mismeasurements of energy depositions in this part of the calorimeter, the performance of the tau identification greatly differs from the rest of the detector.

To remove sensitivity on this effect, events where the leading tau candidate is reconstructed in this region of the calorimeter are rejected. Also events where either of the two leading jets are reconstructed in this region of the calorimeter are rejected, since this would bias the calculation of  $E_T^{\text{miss}}$ .

### 5.2.3 Channel definition

**Rejection of light leptons** To allow for an easy combination with other channels in the future, the analysis explicitly vetoes events with one or more identified isolated electrons or muons. Figure 5.2 shows the number of identified light leptons before any selection criteria are applied.

**Requirement of a second jet** The analysis requires a second jet in the event with  $p_T > 30$  GeV. It is expected in most SUSY scenarios to have multiple high- $p_T$  jets in the event. The requirement on a second jet greatly reduces the background from associated gauge boson + jet production. The multiplicity of jets with  $p_T > 30$  GeV is shown in figure 5.3.

**Requirement of 1 tau lepton** Since this is an inclusive search for supersymmetry with taus in the final states, the analysis requires at least one tau candidate with  $p_T > 15$  GeV in the event.

The identification of tau candidates is done using the boosted decision tree classifier as described in section 4.2. The multiplicity of tau candidates per event after the trigger re-

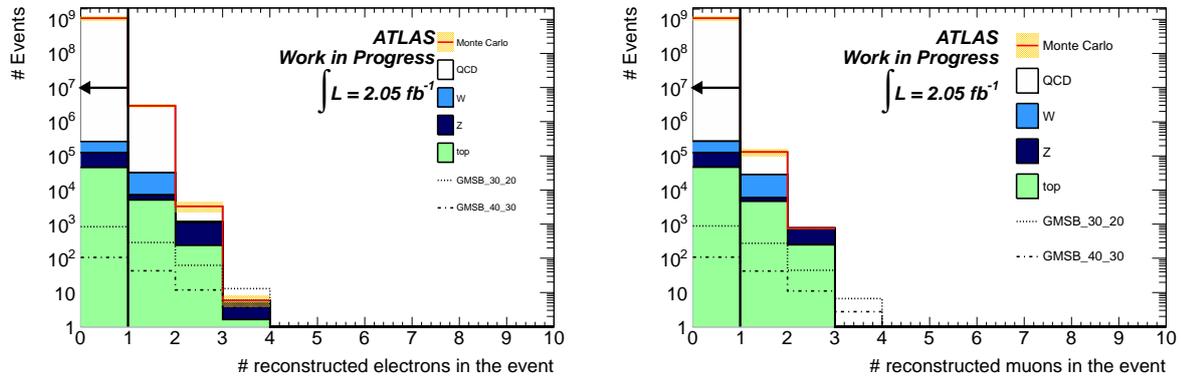


Figure 5.2: Number of identified electrons (left) and muons (right). All plots show the QCD-multijet background in white, the total top background from  $t\bar{t}$  and single top in light green, the total background from Z+jets in dark blue and the total W+jets background in light blue. The red line gives the sum of all Standard Model Samples. If not noted otherwise, the yellow error band gives the uncertainty on the sum of Standard Model backgrounds due to the number of Monte Carlo events.

As an example for a possible signal contribution, two GMSB signals are shown with the dashed and dotted lines.

The vertical lines with an arrow indicate the chosen selection criterion.

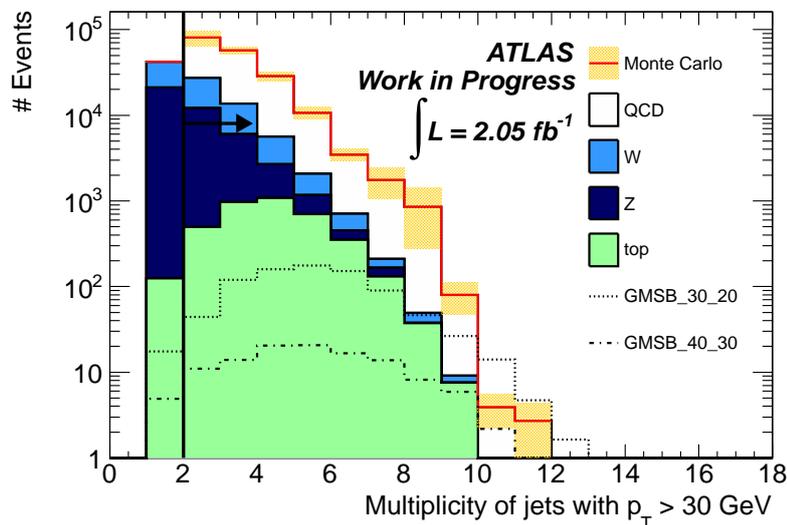


Figure 5.3: Multiplicity of jets with  $p_T > 30 \text{ GeV}$  after the trigger requirements and the rejection of events with identified light leptons.

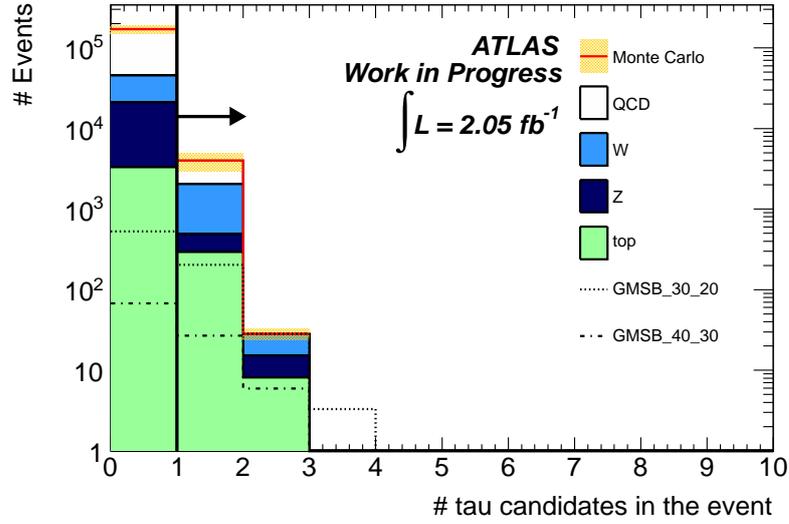


Figure 5.4: Number of tau candidates after all other selection criteria are applied.

requirements, rejection of events with identified light leptons and requiring an additional jet with  $p_T > 30 \text{ GeV}$  is shown in figure 5.4.

#### 5.2.4 Background suppression requirements

The backgrounds considered in this analysis are  $W$ +jets, top production (comprised of  $t\bar{t}$  and single top),  $Z$ +jets and QCD multijet events. The background from diboson production was found to be less than one event in total after the requirement of one tau lepton and is hence neglected for now.

The backgrounds are split up in two classes: Events with real  $E_T^{\text{miss}}$ , e.g. due to neutrinos, and events where the  $E_T^{\text{miss}}$  is caused by detector effects.

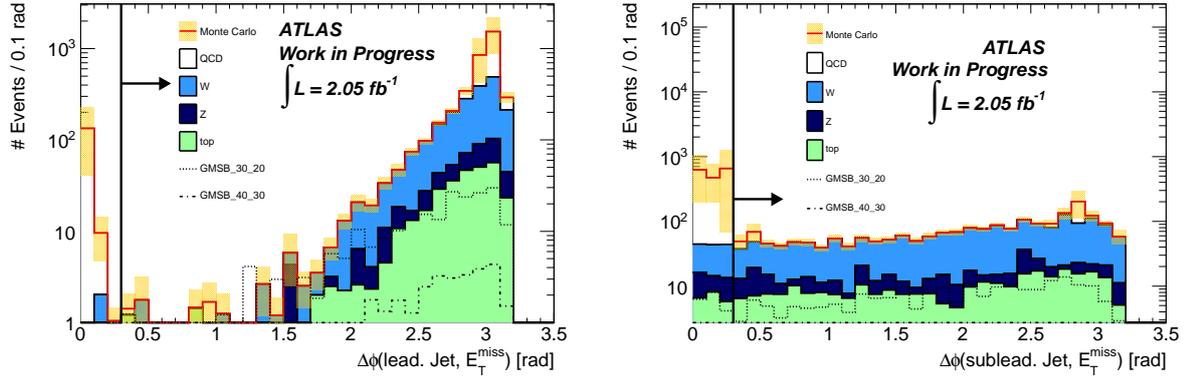
##### Suppression of events with fake $E_T^{\text{miss}}$

The major contribution to the class of events without real  $E_T^{\text{miss}}$  is QCD multijet production where the  $E_T^{\text{miss}}$  is caused by mismeasurement of one or more jets in the event.

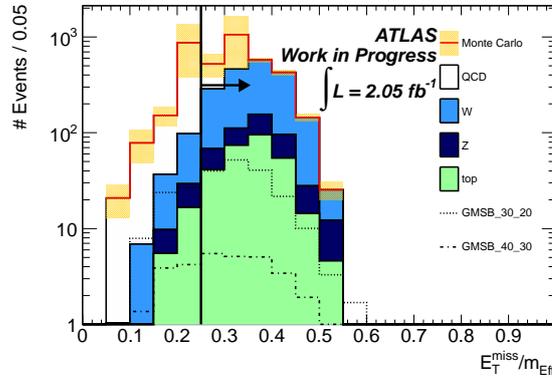
To reject events where the fake  $E_T^{\text{miss}}$  can be associated with a specific mismeasured jet, a requirement on the angular separation  $\Delta\phi$  between this jet and the  $E_T^{\text{miss}}$  is applied. Figure 5.5 shows the distributions of the  $\Delta\phi$  between the  $E_T^{\text{miss}}$  and the leading jet (5.5a) and the  $E_T^{\text{miss}}$  and the subleading jet (5.5b). It can be seen that a large amount of QCD multijet events can be rejected by requiring  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}}) > 0.3 \text{ rad}$ .

If the  $E_T^{\text{miss}}$  cannot be associated with the mismeasurement of one of these jets directly,<sup>2</sup>

<sup>2</sup>The angular correlation between one jet and the  $E_T^{\text{miss}}$  is lost in this case.



(a) Angular separation between  $E_T^{\text{miss}}$  and the leading jet (b) Angular separation between  $E_T^{\text{miss}}$  and the subleading jet



(c)  $E_T^{\text{miss}} / m_{\text{eff}}$

Figure 5.5: Distributions of QCD rejection variables before a requirement on them. The chosen values for the selection criteria are indicated by the vertical line.

the quantity  $E_T^{\text{miss}} / m_{\text{eff}}$  can still show a good separation between events with real  $E_T^{\text{miss}}$  and events where the  $E_T^{\text{miss}}$  is caused by detector effects, since it quantifies the fraction of the total transverse energy in an event carried by  $E_T^{\text{miss}}$ . The analysis requires that 25% of the effective mass of the event is due to  $E_T^{\text{miss}}$ ,  $E_T^{\text{miss}} / m_{\text{eff}} > 0.25$ . Figure 5.5c shows the distribution of this variable before applying the requirement on  $E_T^{\text{miss}} / m_{\text{eff}}$ .

### Suppression of events with real $E_T^{\text{miss}}$

After suppression of QCD multijet events, the major remaining backgrounds are  $W \rightarrow \tau \nu + \text{jets}$ ,  $Z \rightarrow \nu \nu + \text{jets}$  and top production. In  $W$ +jets and top production, events where a tau lepton coming from the decay of a  $W$  boson is correctly identified, as well as events where a jet or lepton is misidentified as a tau lepton contribute. In events where a tau lepton from a  $W$ -decay is correctly identified and the neutrino coming from the same decay, together with

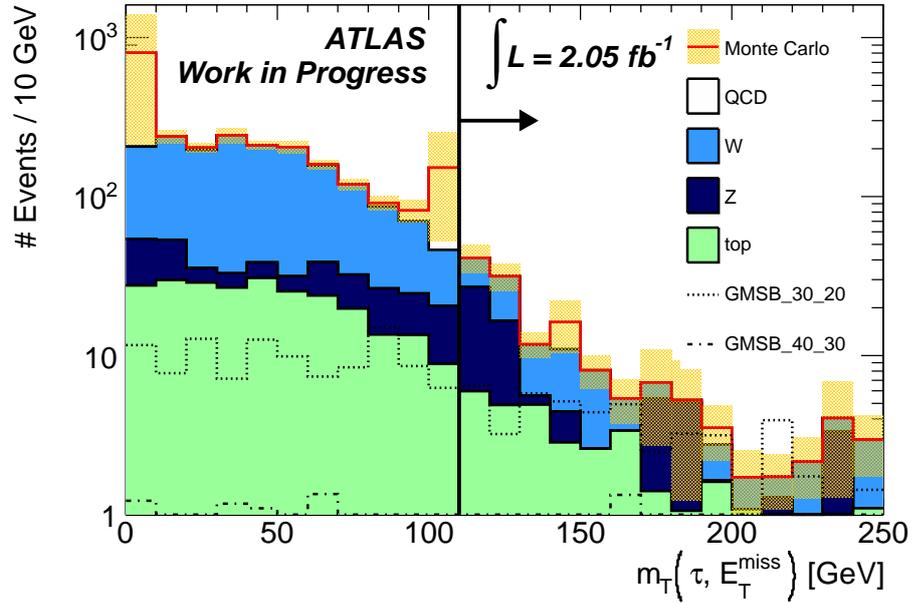


Figure 5.6: Distribution of  $m_T(\tau, E_T^{\text{miss}})$ , shown after the QCD-suppression requirements. The selection criterion is indicated by the black line with an arrow.

the neutrino from the decay of the tau lepton, are the only sources of  $E_T^{\text{miss}}$  in the event, the transverse mass between the selected tau candidate and  $E_T^{\text{miss}}$ ,  $m_T(\tau, E_T^{\text{miss}})$ , has an upper bound at the mass of the  $W$  boson and can be used to suppress a great fraction of  $W$ +jets and top events.

Figure 5.6 shows the distribution of the transverse mass after all other selection criteria are applied. In the analysis the requirement  $m_T(\tau, E_T^{\text{miss}}) > 110 \text{ GeV}$  is chosen.

### Final signal region

After the selection a final discrimination based on the effective mass,  $m_{\text{eff}}$ , is used. The masses of the SUSY particles produced in a collision are expected to be relatively high, the effective mass, which is a measure for the total energy in an event and hence setting the mass scale of the SUSY particles, the signal tends to higher values in  $m_{\text{eff}}$ . Figure 5.7 shows the  $m_{\text{eff}}$  distribution after the requirement on the transverse mass.

In the analysis the requirement  $m_{\text{eff}} > 600 \text{ GeV}$  defines the final signal region SR.

## 5.3 Overview of the full selection

An overview of the event yields at different stages of the selection for the different backgrounds studied is given in table 5.3.

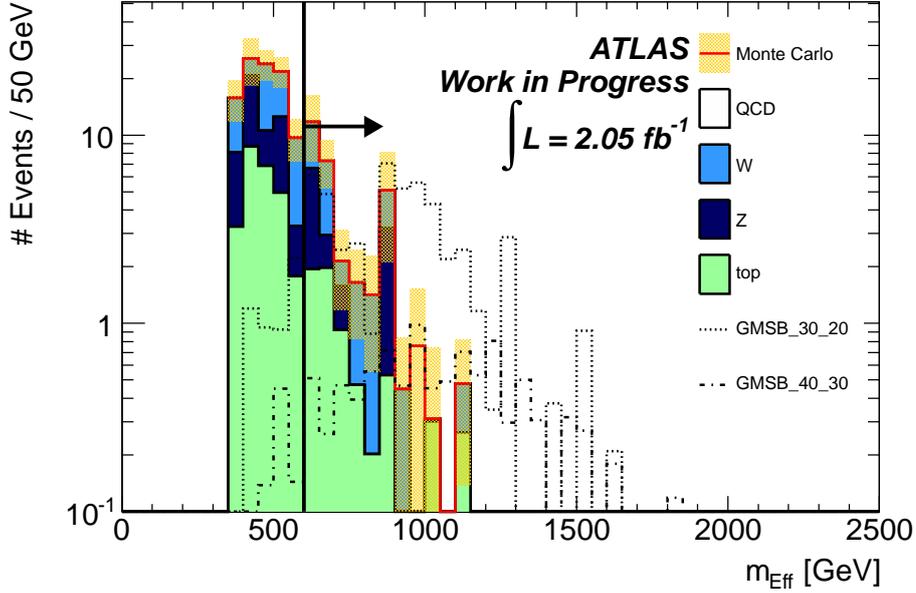


Figure 5.7: Distribution of the effective mass  $m_{\text{eff}}$ , shown after all other selection criteria are applied.

The event numbers after the final requirement can be compared to the Monte Carlo expectations for the two different GMSB benchmark points, shown in the figures:

$$N_{\text{GMSB}_{\Lambda=30}^{\tan\beta=20}} = (49.8 \pm 5.1) \text{ events} \quad \text{and}$$

$$N_{\text{GMSB}_{\Lambda=40}^{\tan\beta=30}} = (9.7 \pm 1.0) \text{ events} .$$

This Monte Carlo based overview of the selection shows, that contributions from  $W$ +jets, top and  $Z$ +jets production are expected to be larger than the contribution from QCD multijet events. It also shows, that the Monte Carlo estimate of the QCD multijet production cannot be trusted because of the limited statistical precision of the QCD multijet samples after the channel definition, which is mostly due to the requirement of an identified tau lepton in the event. As a first attempt to increase the statistical precision of the Monte Carlo estimate for the QCD multijet samples, a Monte Carlo event weighting method is applied, described briefly in the following section.

|  | top              | W+jets                         | Z+jets                         | QCD                            |
|--|------------------|--------------------------------|--------------------------------|--------------------------------|
| Before selection                                 | $314910 \pm 390$ | $(6432.4 \pm 3.2) \times 10^4$ | $(8519.9 \pm 3.5) \times 10^3$ | $(4669.2 \pm 4.9) \times 10^6$ |
| Preselection                                     | $300 \pm 12$     | $1556 \pm 55$                  | $230 \pm 11$                   | $660 \pm 280$                  |
| channel definition                               | $300 \pm 12$     | $1556 \pm 55$                  | $230 \pm 11$                   | $660 \pm 280$                  |
| QCD suppression                                  | $267.7 \pm 8.9$  | $1379 \pm 44$                  | $181.2 \pm 7.5$                | $3.1 \pm 2.8$                  |
| $m_T(\tau, E_T^{\text{miss}}) > 110 \text{ GeV}$ | $32.3 \pm 3.1$   | $56.3 \pm 5.6$                 | $41.9 \pm 2.4$                 | $0.3 \pm 0.3$                  |
| $m_{\text{eff}} > 600 \text{ GeV}$               | $6.7 \pm 1.4$    | $15.0 \pm 2.9$                 | $11.8 \pm 1.2$                 | $0.3 \pm 0.3$                  |

Table 5.3: Expected numbers of events from Monte Carlo simulation, normalized to  $2.05 \text{ fb}^{-1}$ . No normalization factors are included. The uncertainties given are statistical only.

### 5.3.1 QCD event weighting using tau misidentification probabilities

Table 5.3 shows that a large part of the QCD multijet events is rejected when asking for an identified tau lepton. To increase the statistical precision, this method uses misidentification probabilities of jets faking a tau lepton and weighting the Monte Carlo events with these probabilities instead of completely rejecting events without an identified tau candidate.

The tau misidentification probability is obtained by performing a completely separate event selection. A sample of *dijet* events is selected by requiring two reconstructed jets which are balanced in  $p_T$  and “back-to-back” in  $\phi$ . To get events over a large range of jet  $p_T$ , a low-threshold single jet trigger is used and matched to the leading jet (“tag jet”). The subleading jet is required to have at least one associated track, since the tau identification relies on track variables that are not well defined for tau candidates without tracks. This jet is then used to probe the tau misidentification probability (“probe jet”). The misidentification probability  $f_{ID}$  is then calculated as:

$$f_{ID} = \frac{\# \text{ of probed jets identified as tau leptons}}{\# \text{ of all reconstructed probe jets}}. \quad (5.1)$$

The numerator in this expression only includes probe jets that satisfy all identification criteria from section 4.2. A detailed discussion of the dijet selection and calculation of the tau misidentification probabilities can be found in [37].

Since the tau misidentification probability is known to be very dependent on the  $p_T$  of the tau candidate, this misidentification probability is computed in bins of tau candidate<sup>3</sup>  $p_T$ . Figure 5.8 shows the misidentification probabilities obtained with this algorithm versus the probe-tau candidate  $p_T$ .

<sup>3</sup>Tau candidate in this context means a reconstructed jet with at least one track associated with it.

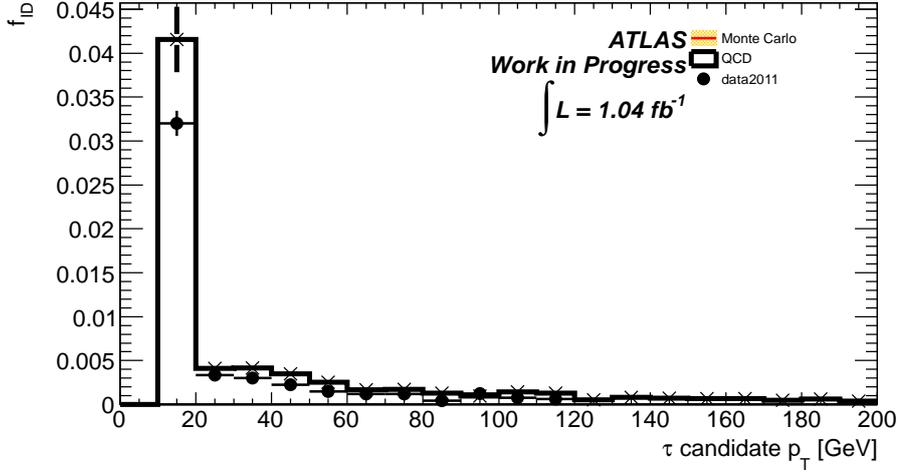


Figure 5.8: Tau misidentification probabilities for the used  $BDT_{\tau}$  working point obtained from a dijet sample with a tag-and-probe method. The black lines show the misidentification probability in QCD multijet Monte Carlo and the black dots are obtained from data. Taken from [32].

To turn the misidentification probabilities into event weights, they are applied to every tau candidate in simulated QCD events. If there are  $n$  tau candidates in an event all combinations of tau identification pass and fail decision are considered. The total probability  $w$  of an event using the measured misidentification probability  $f_{ID,i}$  for each of the  $n$  tau candidates is then defined as:

$$w = \prod_{i=1}^n x_i \quad \text{with } x_i = \begin{cases} f_{ID,i} & \text{if candidate } i \text{ passes ,} \\ (1 - f_{ID,i}) & \text{if candidate } i \text{ fails .} \end{cases} \quad (5.2)$$

From all combinations with at least one tau candidate that passes the identification criteria, one is chosen randomly. After selecting one of the tau candidates, the event is propagated through the event selection with a weight corresponding to its total probability  $w$ . The effect of this method on the QCD multijet event yield is shown in table 5.4.

The results obtained with this method are used as a cross-check for the number of QCD multijet events in both the signal region and in a control region, described in section 6.1.1.

In the following chapters the estimation of the  $W$ +jets, top and QCD-multijet backgrounds, based on data, are presented.

|  | nominal QCD   | reweighted QCD |
|--|---------------|----------------|
| channel definition   | $660 \pm 280$ | $760 \pm 350$  |
| QCD suppression  | $3.1 \pm 2.8$ | $47 \pm 22$    |
| $m_{\text{T}}(\tau, E_{\text{T}}^{\text{miss}}) > 110 \text{ GeV}$ | $0.3 \pm 0.3$ | $3.8 \pm 1.8$  |
| $m_{\text{eff}} > 600 \text{ GeV}$                                 | $0.3 \pm 0.3$ | $0.4 \pm 0.1$  |

Table 5.4: Comparison of the QCD multijet event yields for the Monte Carlo samples with and without tau misidentification probability reweighting throughout the selection.

# Chapter 6 Estimation of the $W$ +jets and top backgrounds

---

In this section, the techniques developed for estimating the backgrounds from  $W$ +jets and top are described.

The Monte Carlo-based estimates for the selection, summarized in table 5.3, show that the  $W$ +jets and top backgrounds are important backgrounds for this analysis. This is due to the fact that both of these backgrounds can have tau leptons and genuine  $E_T^{\text{miss}}$  in the event from the decay  $W \rightarrow \tau\nu$ , either from direct  $W$ +jets production or the decay of a top quark.

To estimate the  $W$ +jets and top backgrounds, two control regions are defined, one for events with true tau leptons, the second for events with misidentified tau candidates. In the true-tau control region  $CR_1$ , the contributions from  $W$ +jets and top processes are separated, and the event yield for each of the backgrounds in data is compared with the Monte Carlo expectation. *Two separate scale factors* are determined that account for the normalization difference between data and Monte Carlo for the two different samples.

In the second control region  $CR_2$ , that has also sizeable contributions from events with misidentified tau leptons, a *combined scale factor* for events with misidentified tau candidates is determined.

The  $W$ +jets and top backgrounds can be split up in two different classes based on simulation: Events where a real tau lepton from a  $W$ -decay is selected in the analysis and there is no mismeasurement of the  $E_T^{\text{miss}}$  and events where a misidentified tau candidate is selected. Figure 6.1 shows the distributions of  $m_T(\tau, E_T^{\text{miss}})$  for events with truth-matched tau leptons and events with misidentified tau leptons before the requirements on  $m_T(\tau, E_T^{\text{miss}})$  and  $m_{\text{eff}}$ .

Since the rates of the two different classes of events could be modeled differently well by simulation, two different control regions need to be defined to obtain separate scale factors for the different classes of events.

As can be seen from figure 6.1, events with real tau leptons are mainly located at low  $m_T(\tau, E_T^{\text{miss}})$ , whereas events with misidentified tau candidates are more spread out to

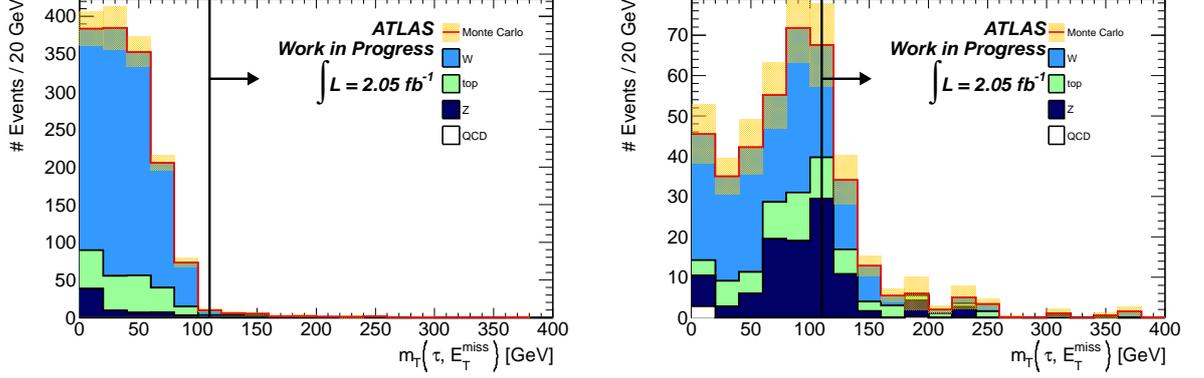


Figure 6.1:  $m_T(\tau, E_T^{\text{miss}})$  for events with identified (left) and misidentified tau candidates (right). The black lines denote the chosen selection criterion for the selection.

|                 | Data | W+jets        | top             | Z+jets         | QCD           |
|-----------------|------|---------------|-----------------|----------------|---------------|
| $N^{\text{CR}}$ | 1256 | $1118 \pm 42$ | $185.0 \pm 7.2$ | $85.2 \pm 9.1$ | $7.7 \pm 7.5$ |

Table 6.1: Number of events for data and background simulation in the low- $m_T(\tau, E_T^{\text{miss}})$  control region  $\text{CR}_1$ .

higher  $m_T(\tau, E_T^{\text{miss}})$ .

Thus a true tau enriched control region  $\text{CR}_1$  is defined with the requirement  $m_T(\tau, E_T^{\text{miss}}) < 70 \text{ GeV}$ . In this control region the purity of events with real tau leptons is 93.2% for W+jets and top and it is thus an adequate control region to obtain the scale factor for events with real tau leptons.

Figure 6.2 shows the definitions of the true-tau control region  $\text{CR}_1$  and the final signal region, SR, in the  $((m_T(\tau, E_T^{\text{miss}}) - m_{\text{eff}})$ -plane. The remaining region in the  $((m_T(\tau, E_T^{\text{miss}}) - m_{\text{eff}})$ -plane is used as a control region to obtain the scale factor for events with misidentified tau candidates,  $\text{CR}_2$ .

## 6.1 W+jets / top background with real tau leptons

To estimate the W+jets and top backgrounds in the analysis, the true-tau enhanced control region  $\text{CR}_1$  is used. Table 6.1 shows the number of events in data and for the different background Monte Carlo samples in this control region.

Before estimating the W+jets and top, a data-driven cross-check of the number of QCD-

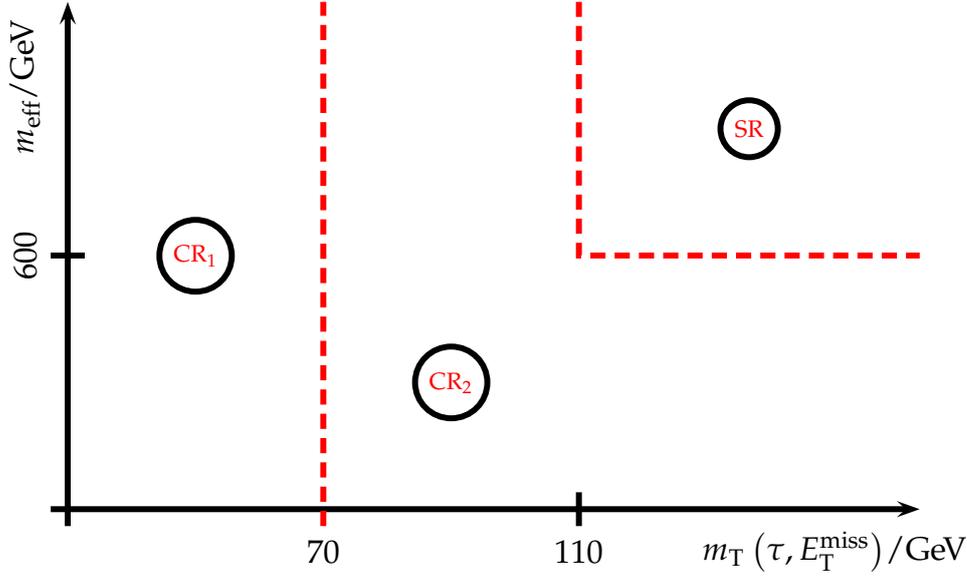


Figure 6.2: Definition of the two control regions (CR<sub>1</sub> & CR<sub>2</sub>) and the signal region (SR) in the  $((m_T(\tau, E_T^{\text{miss}}) - m_{\text{eff}})$ -plane.

multijet events in CR<sub>1</sub> obtained by reweighting the MC as explained in section 5.3.1, is carried out.

### 6.1.1 Estimation of the QCD background in CR<sub>1</sub>

The strategy of the data-driven method to estimate the normalization of the QCD background in CR<sub>1</sub> is to define two different QCD enhanced control regions. The first one, QCD<sub>1</sub>, is obtained by applying all CR<sub>1</sub> selection criteria, including the requirement for an identified tau lepton and the requirement on  $m_T(\tau, E_T^{\text{miss}})$ , but inverting the QCD rejection requirements  $\Delta\phi(\text{jet1}, E_T^{\text{miss}})$  and  $\Delta\phi(\text{jet2}, E_T^{\text{miss}})$  (for simplicity the combination of these two requirements are now referred as  $\Delta\phi$  symbolically) and  $E_T^{\text{miss}}/m_{\text{eff}}$  simultaneously. This ensures that the control region is as close as possible to CR<sub>1</sub> in phase space, but still has a sizeable amount of QCD background left. The second control region, QCD<sub>2</sub>, is defined by inverting the requirement of an identified tau candidate, thus selecting events with zero tau candidates and omitting all subsequent steps in the selection.

Tables 6.2 and 6.3 show the number of events for data and Monte Carlo samples in the two QCD enhanced regions. Both regions are clearly dominated by QCD-multijet events and have a fair amount of events left.

The number of QCD events in CR<sub>1</sub> is then obtained by subtracting off the non-QCD contributions from data using simulation in QCD<sub>1</sub> and applying transfer factors to estimate the number of QCD-multijet events in CR<sub>1</sub>.

|  | data | QCD           | Z+jets          | top             | W+jets          |
|--|------|---------------|-----------------|-----------------|-----------------|
| $n_\tau > 0$                                     | 2308 | $780 \pm 360$ | $229 \pm 11$    | $299 \pm 12$    | $1548 \pm 1548$ |
| $\Delta\phi$ requirements                        | 2211 | $770 \pm 360$ | $220.1 \pm 9.2$ | $304.3 \pm 9.6$ | $1544 \pm 44$   |
| $E_T^{\text{miss}}/m_{\text{eff}} > 0.25$        | 181  | $175 \pm 24$  | $5.0 \pm 1.5$   | $5.2 \pm 1.4$   | $18.0 \pm 3.1$  |
| $m_T(\tau, E_T^{\text{miss}}) > 110 \text{ GeV}$ | 90   | $77 \pm 15$   | $4.2 \pm 1.5$   | $3.7 \pm 1.2$   | $11.8 \pm 2.5$  |

Table 6.2: Numbers of observed and expected events in  $\text{QCD}_1$ .

|              | data  | top           | W+jets          | Z+jets          | QCD                          |
|--------------|-------|---------------|-----------------|-----------------|------------------------------|
| $n_\tau = 0$ | 90285 | $3335 \pm 32$ | $24350 \pm 140$ | $18010 \pm 260$ | $(45.4 \pm 4.2) \times 10^3$ |

Table 6.3: Numbers of observed and expected events in the zero-tau region  $\text{QCD}_2$ .

$$\hat{N}^{\text{QCD}} = \left( N_{\text{QCD}_1}^{\text{data}} - N_{\text{QCD}_1}^{\text{non-QCD}} \right) \cdot t_{\Delta\phi} \cdot t_{E_T^{\text{miss}}/m_{\text{eff}}} \quad (6.1)$$

The second region,  $\text{QCD}_2$  is then used to factorize the two requirements on  $\Delta\phi$  and  $E_T^{\text{miss}}/m_{\text{eff}}$ . The transfer factors  $t_{\Delta\phi}$  and  $t_{E_T^{\text{miss}}/m_{\text{eff}}}$  from equation 6.1 are derived from the  $\text{QCD}_2$  zero-tau sideband by applying the requirement on  $\Delta\phi$  to obtain the transfer factor from the low to the high  $E_T^{\text{miss}}/m_{\text{eff}}$  region and vice-versa. Again this is done by subtracting off the remaining non-QCD backgrounds from data, which is only a small correction since the purity of QCD events after each of the two QCD-rejection requirements is still expected to be larger than 90%.

$$t_{\Delta\phi} \equiv \left( \frac{N_{\text{QCD}_2}^{\text{data, pass } \Delta\phi} - N_{\text{QCD}_2}^{\text{non-QCD, pass } \Delta\phi}}{N_{\text{QCD}_2}^{\text{data, fail } \Delta\phi} - N_{\text{QCD}_2}^{\text{non-QCD, fail } \Delta\phi}} \right)_{E_T^{\text{miss}}/m_{\text{eff}} < 0.25}, \quad (6.2)$$

and similarly for the second transfer factor  $t_{E_T^{\text{miss}}/m_{\text{eff}}}$ :

$$t_{E_T^{\text{miss}}/m_{\text{eff}}} \equiv \left( \frac{N_{\text{QCD}_2}^{\text{data, pass } E_T^{\text{miss}}/m_{\text{eff}}} - N_{\text{QCD}_2}^{\text{non-QCD, pass } E_T^{\text{miss}}/m_{\text{eff}}}}{N_{\text{QCD}_2}^{\text{data, fail } E_T^{\text{miss}}/m_{\text{eff}}} - N_{\text{QCD}_2}^{\text{non-QCD, fail } E_T^{\text{miss}}/m_{\text{eff}}}} \right)_{\Delta\phi < 0.4}. \quad (6.3)$$

The scale factors derived by this method are:

$$t_{\Delta\phi} = 0.1 \text{ and} \quad (6.4)$$

$$t_{E_T^{\text{miss}}/m_{\text{eff}}} = 0.85 . \quad (6.5)$$

This technique to obtain an estimate for the number of QCD events in  $\text{CR}_1$  only works under the assumption that all variables used ( $\Delta\phi$ ,  $E_T^{\text{miss}}/m_{\text{eff}}$  and  $n_\tau$ ) are uncorrelated. This assumption is tested by looking at 2-dimensional projections of these variables for the different samples in Monte Carlo and data.

It is found that there is a correlation between the angular separation of the  $E_T^{\text{miss}}$  to the leading jets and  $E_T^{\text{miss}}/m_{\text{eff}}$  for the non-QCD backgrounds ( $W$ +jets,  $Z$ +jets and top). Considering the small contribution of these backgrounds with respect to the number of QCD multijet events in  $\text{QCD}_2$ , this correlation is estimated to be negligible for the extrapolation from  $\text{QCD}_1$  to  $\text{CR}_1$ . In the QCD multijet samples no correlation between  $\Delta\phi$  and  $E_T^{\text{miss}}$  is seen. This expectation is met in data. Figure 6.3 shows the number of events in the  $(E_T^{\text{miss}}/m_{\text{eff}}, \Delta\phi(E_T^{\text{miss}}, \text{jet}_{1,2}))$ -plane for the  $W$ +jets, top and QCD-multijet samples from simulations and from data.

For the QCD background a correlation between the number of tau candidates in the event and  $\Delta\phi$  is found. This can be explained as a combinatorial effect; events with less jets are inherently less likely to have a misidentified tau candidate from one of the jets and simultaneously less likely to have a poorly reconstructed jet accounting for the  $E_T^{\text{miss}}$ . To estimate this effect on the transfer factors from  $\text{QCD}_1$  into  $\text{CR}_1$ , the Monte Carlo sample is split up into three subsamples:

- 2 reconstructed jets, 0 tau candidates
- At least 3 reconstructed jets, 0 tau candidates
- 2 reconstructed jets, at least 1 tau candidate.

The  $\Delta\phi(E_T^{\text{miss}}, \text{jet}_{1,2})$  distributions for the three subsamples are shown for the QCD-multijet simulation in figure 6.4.

In each of these subsamples the two transfer factors are determined independently. The results for the transfer factors vary by about 50% with a statistical uncertainty of about 50%. The transfer factors are summarized in table 6.4. To account for the correlations a conservative 100% uncertainty on the combined transfer factor from  $\text{QCD}_1$  to  $\text{CR}_1$  is assumed.

The total estimate of the QCD background in  $\text{CR}_1$  is:

$$\hat{N}^{\text{QCD}} = (7.7 \pm 7.7) \text{ events.} \quad (6.6)$$

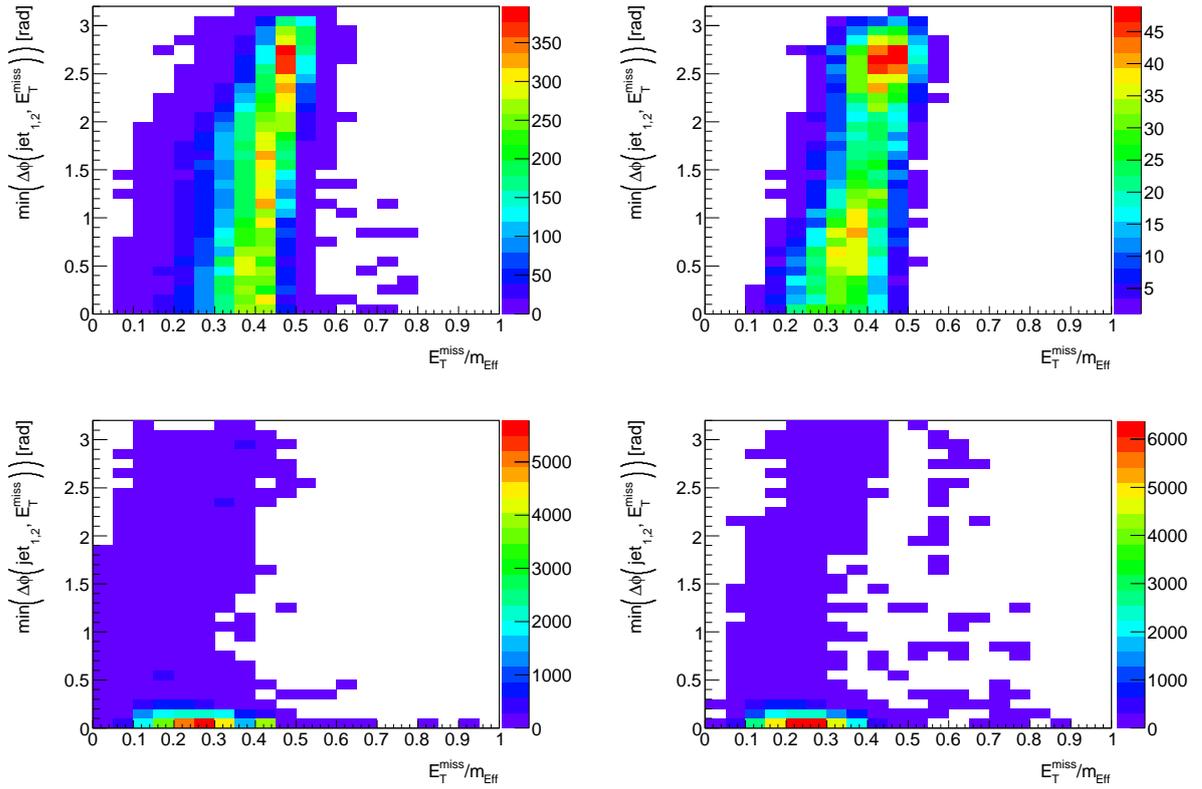


Figure 6.3: Number of events in the  $(E_T^{\text{miss}}/m_{\text{eff}}, \Delta\phi(E_T^{\text{miss}}, \text{jet}_{1,2}))$ -plane for the  $W$ +jets (top left), top (top right) and QCD-multijet (bottom left) samples from simulations and from data (bottom right).

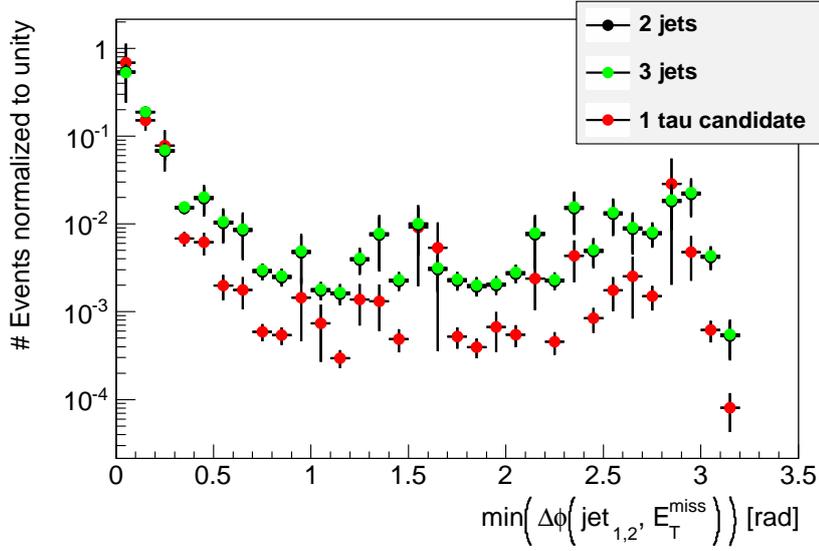


Figure 6.4:  $\Delta\phi(E_T^{\text{miss}}, \text{jet}_{1,2})$  distribution for the three subsamples defined in section 6.1.1. The distributions are normalized to unity. The shapes of the first two subsamples are almost identical, thus the markers of the first subsample are hidden behind the markers of the second.

| Subsample       | $t_{\Delta\phi}$ | $t_{E_T^{\text{miss}}/m_{\text{eff}}}$ |
|-----------------|------------------|--|
| 2 jets          | $0.37 \pm 0.19$  | $0.9 \pm 0.3$                          |
| 3 jets          | $0.39 \pm 0.19$  | $0.9 \pm 0.3$                          |
| 1 tau candidate | $0.2 \pm 0.6$    | $2.9 \pm 0.6$                          |

Table 6.4: Transfer factors derived in the three subsamples defined in section 6.1.1.

This estimate is consistent with the one obtained with the fake misidentification probability method, which is  $(7.7 \pm 7.5)$  events. The shape of the QCD background distributions is taken from the Monte Carlo sample, reweighted with the misidentification probabilities and the normalization from this data-driven method.

### 6.1.2 Separation of $W$ +jets from top events

Instead of deriving a combined scale factor for the  $W$ +jets and top contributions in  $\text{CR}_1$  an attempt is made to estimate the two contributions separately. The idea behind this is that, even though the comparison between data and Monte Carlo in figure 6.5 shows good agreement, there is the possibility that their relative contributions are not predicted correctly since the  $W$ +jets and top samples are simulated with different event-generators using different

underlying physics models (see section 5.1).

The strategy to separate  $W$ +jets from top events is to find different kinematic or event-shape variables that can be used to classify the events.

A large number of different variables have been studied as input to the classification of  $W$ +jets and top events. The following four variables show the best separation power among these:

- The **number of reconstructed  $b$ -jets** in the event.  
This variable showed the best discrimination power. The top quark predominantly decays to a  $W$  boson and a  $b$ -quark (see figure 6.6) and the majority of the top events in the signal region come from  $t\bar{t}$ -decays. The efficiency of the used  $b$ -tagger at the used working point is 60% (see chapter 4). In  $CR_1$  65.3% of all top events but only 3.9% of all  $W$ +jets events have at least one tagged  $b$ -jet with  $p_T > 20$  GeV.
- The **jet multiplicity** of the event.  
In  $W \rightarrow \tau\nu_\tau$  events, where the tau is reconstructed correctly, additional jets occur in the production process of the  $W$  boson (see figure 6.7). In  $t\bar{t}$  events in addition to the 2  $b$ -jets, two more jets can be produced from a hadronic decay of the second  $W$  boson.
- The **transverse thrust  $T_T$**  of the event (see section 4.7).  
Aside from  $E_T^{\text{miss}}$  caused by neutrinos or mismeasured objects, events have to be balanced in  $p_T$ , which means that events with more objects look more spherical (small transverse thrust) and events with fewer objects appear more “pencil-like” (large transverse thrust). Thus top events tend to lower values of  $T_T$  compared to  $W$ +jets events.
- The  **$p_T$  of the second leading jet**.  
The  $p_T$  spectrum of jets produced in association with a  $W$  boson falls steeply above the selection threshold of  $p_T > 30$  GeV. On the other hand the jets involved in decays of top quarks are more energetic due to the large mass of the top quark. This leads to a harder jet  $p_T$  spectrum in top events compared to  $W$ +jets events. Due to the harsh trigger requirement on the leading jet, the  $p_T$  of the subleading jet is used as a separation variable.

The distributions of these four variables in  $CR_1$  are shown in figure 6.8. What strikes the eye when comparing data with Monte Carlo for these variables, is the discrepancy in the number of reconstructed  $b$ -jets. The Monte Carlo clearly tends to lower values with respect to the data. Also the other variables show clear discrepancies. Because of these discrepancies the contributions from  $W$ +jets and top are scaled separately to improve the agreement.

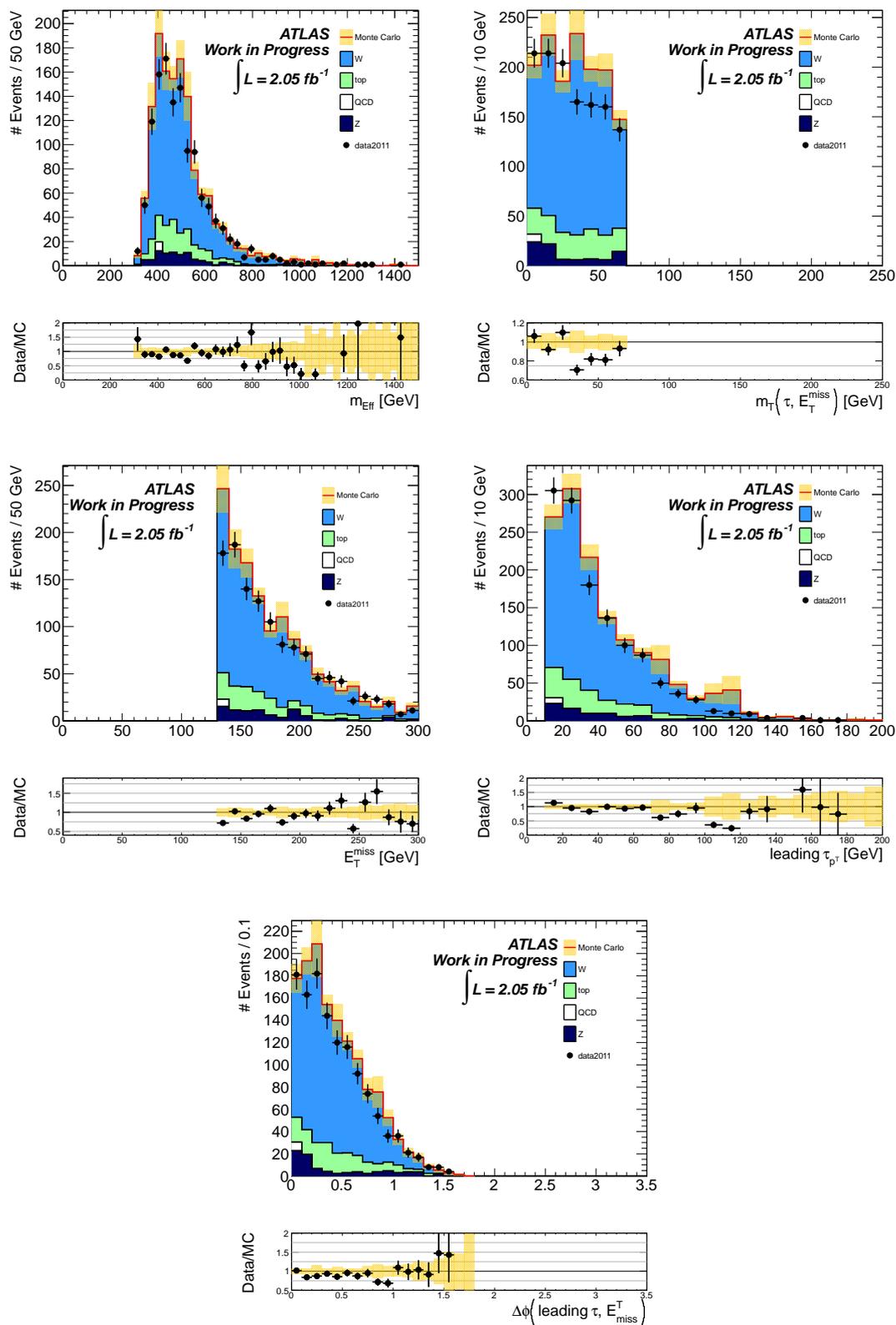


Figure 6.5: Kinematic distributions in the true-tau control region  $CR_1$ . The errors shown are statistical only.

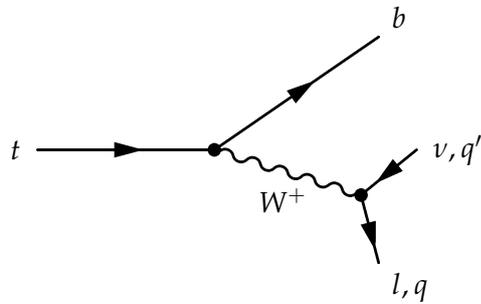


Figure 6.6: Feynman diagram for the decay of the top quark.

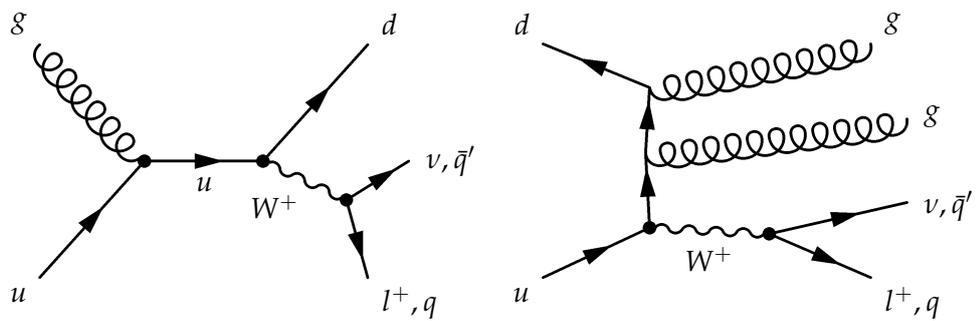


Figure 6.7: Example feynman diagrams for the production of a  $W$  boson in association with 1 jet (left) and 2 jets (right).

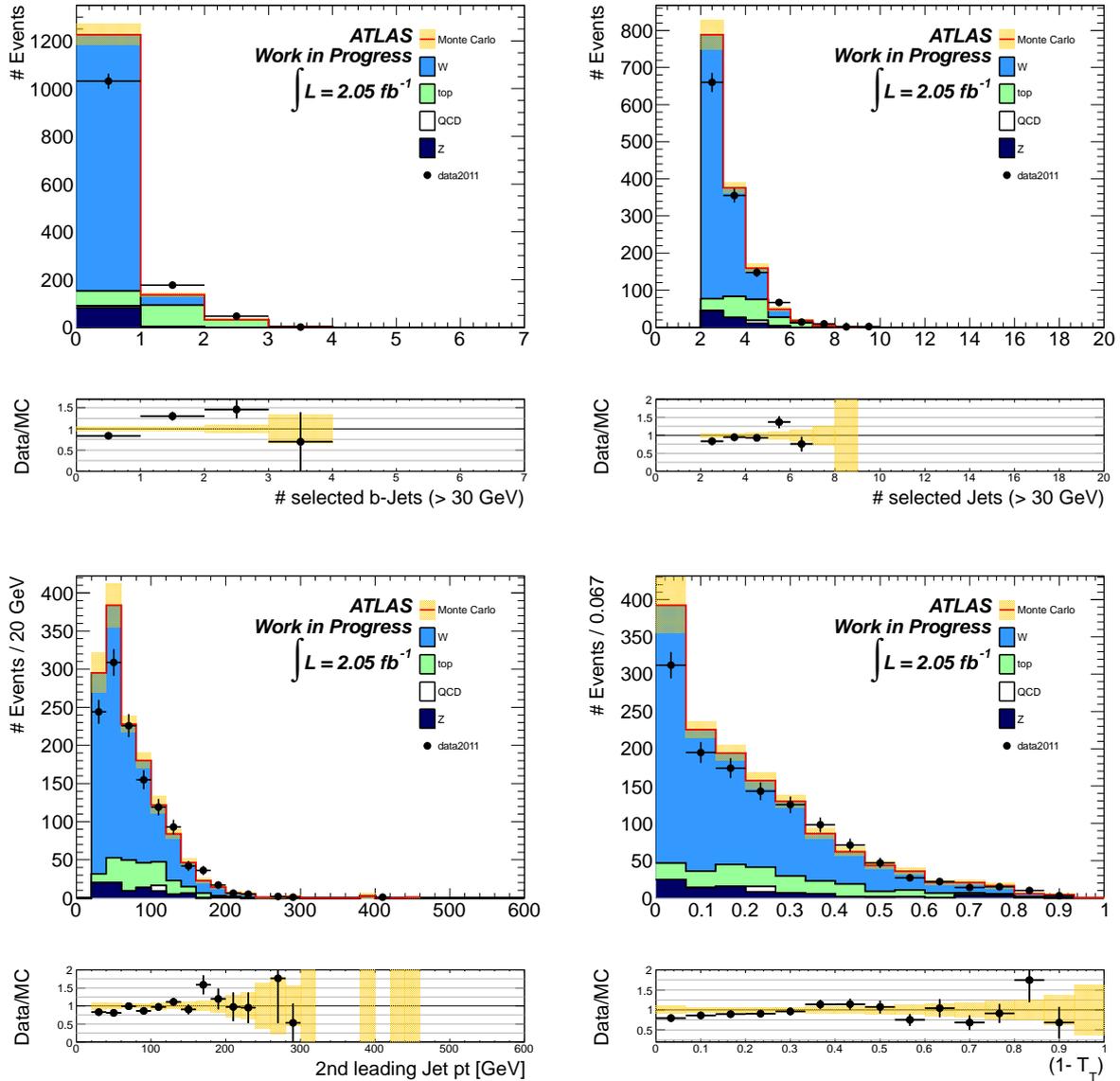


Figure 6.8: Input variables to the classification of  $W$ +jets and top events, shown in the true-tau control region  $CR_1$ . The number of  $b$ -jets (upper left), the jet multiplicity (upper right), the  $p_T$  of the subleading jet (lower right) and  $(1 - T_T)$  (lower right). The errors on data and the sum of all Monte Carlo samples are statistical only.

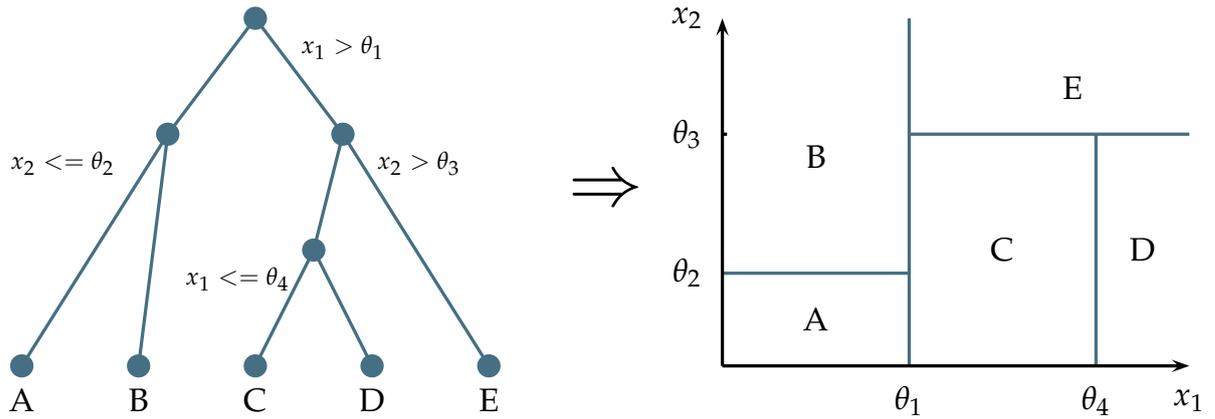


Figure 6.9: Schematic view of the classification of events using a Decision Tree (left) and the corresponding regions of the input parameter space (right). Adapted from [38].

Instead of a simple multi-dimensional cut in the space that is spanned by these variables, a different multivariate technique is used for the classification. A multivariate method classifies events based on more than one input variable. The easiest of these methods would be the rectangular cut method mentioned above. In this analysis, a so-called **Boosted Decision Tree** is trained and used for the classification of the events.

### General remarks about Boosted Decision Trees

A Decision tree partitions the space of the input variable into rectangles by repeated yes/no decisions. These “leaves” represent different regions in the parameter space and are classified either signal or background according to the majority of training events in this leaf.

To use a decision tree on data the classifier has to be trained beforehand. It uses the separation variables in the different Monte Carlo samples and assigns every event to a final leaf as explained above. An example for illustration of a BDT with just two input values  $x_1$  and  $x_2$  is shown in figure 6.9.

One problem of this approach is the sensitivity to statistical fluctuations in the input samples that can bias the classification. Another drawback is that with this method it is very hard to capture complex structures in the input distributions.

A solution to these problems comes in the form of “boosting”. Instead of doing the classification only once, the method of the Boosted Decision Tree (BDT) trains a large number of subsequent trees (a “forest”) and increases the weights of misclassified events. After a fixed number of trees is reached, each event is classified according to the majority of single tree votes.

The training of the BDT used in this analysis was done with TMVA [39], a toolkit specific-

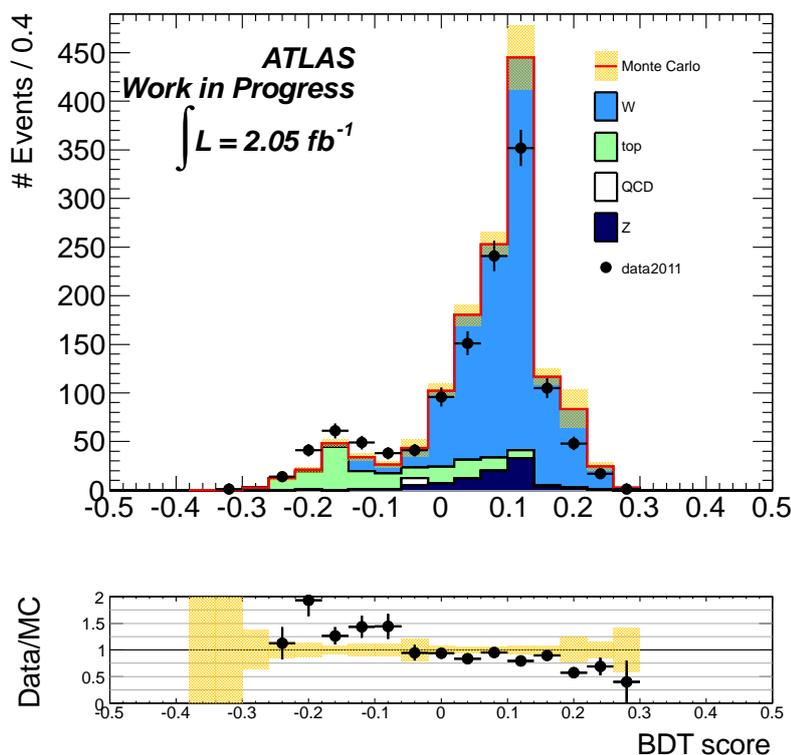


Figure 6.10: BDT discriminant computed from the input variables in figure 6.8 in  $\text{CR}_1$ .

ally designed for multivariate analysis. The package takes the distributions of the chosen separation variables for signal and background separately as inputs and splits these samples randomly into a “training” and a “test” sample. The training sample is used to train the classifier whose output is then used on the statistically independent test sample.

The four variables are used to train the  $W$ /top separation BDT. Only the  $W$ +jets (signal) and top samples (background) are used for the training process. Figure 6.10 shows the resulting BDT distribution evaluated at all Monte Carlo sample and data. From this distribution it can clearly be seen that the Monte Carlo simulation predicts too many  $W$ +jets and too few top events. This leads to a scale factor smaller than 1 for  $W$ +jets and larger than 1 for the top contribution.

The computed BDT shows a good separation between  $W$ +jets and top events. This discriminant is used to determine the relative fractions of the  $W$ +jets and top contributions by fitting the BDT distribution in data using a binned, extended maximum-likelihood fit, the Barlow-Beeston fit algorithm. This fit algorithm takes into account the statistical uncer-

tainty on Monte Carlo samples (here  $W$ +jets and top). The concept of the fitting method is described in the following paragraph, a detailed discussion of the fit can be found in the original publication by Barlow and Beeston [40].

### Barlow-Beeston Fit

For a given data sample and  $m$  different Monte Carlo sources, binned in  $n$  bins,  $d_i$  is the number of data events and  $X_i^m$  is the number of simulated events of the  $m^{\text{th}}$  Monte Carlo source in bin  $i$ . The sum of all Monte Carlo sources in the  $i^{\text{th}}$  bin,  $X_i$  can then be written as

$$X_i = \sum_{j=1}^m p_j \cdot a_{ij} \quad (6.7)$$

where  $a_{ij}$  is the number of simulated events for Monte Carlo source  $j$  in bin number  $i$  and  $p_j$  is the fraction of the  $j^{\text{th}}$  Monte Carlo source with respect to the sum of all sources in the fit. These  $p_j$  are exactly the quantities that need to be determined in the fit.

If the number of data events in every bin were large enough so that they can be treated as normally distributed, the  $p_j$  can be estimated by minimizing the  $\chi^2$  defined as

$$\chi^2 = \sum_i \frac{(d_i - X_i)^2}{d_i}. \quad (6.8)$$

As can be seen from figure 6.10, the number of data events in each bin cannot be treated as Gaussian distributed due to the limited number of events. Instead of using a Gaussian approximation, one needs to use the Poisson probability of observing  $d_i$  data events in bin  $i$ :

$$e^{-X_i} \frac{X_i^{d_i}}{d_i!} \quad (6.9)$$

and estimate the  $p_j$  by maximizing the logarithm of the total likelihood

$$\ln \mathcal{L} = \sum_{i=1}^n d_i \ln X_i - X_i, \quad (6.10)$$

which accounts correctly for small numbers of data events in each bin.

The extension of this method that also takes into account the fact that not only the number of data events in some bins might be small, but also the limited number of simulated Monte Carlo events, is the Barlow-Beeston fit method.

The modification that has to be done is to substitute the generated number of events in each bin,  $a_{ij}$ , with the expected number of events  $A_{ij}$  in each bin (which are unknown) in equation (6.7) and to maximize the combined likelihood of the observed  $d_i$  and the observed  $a_{ij}$

$$\ln \mathcal{L} = \sum_{i=1}^n \frac{d_i A_{ij}}{X_i} - A_{ij} \quad (6.11)$$

To maximize this likelihood the derivative is taken and set to zero, which leads to two coupled sets of equations:

$$\sum_{i=1}^n \frac{d_i A_{ij}}{X_i} - A_{ij} = 0 \quad \forall j \text{ and} \quad (6.12)$$

$$\frac{d_i p_j}{X_i} - p_j + \frac{a_{ij}}{A_{ij}} - 1 = 0 \quad \forall i, j. \quad (6.13)$$

The second equation can be simplified to

$$A_{ij} = \frac{a_{ij}}{1 + p_j \left(1 - \frac{d_i}{X_i}\right)}. \quad (6.14)$$

Using a given set of  $p_j$  as starting values, the expected values of  $A_{ij}$  can be calculated and inserted into equation 6.12, which can then be maximized to obtain the best values for the  $p_j$ .

The Barlow-Beeston fit method is provided as a class of the ROOT [41] analysis framework called `TFractionFitter`, which is used to perform the fit in this analysis. The implementation of the fit has a shortcoming which makes it impossible to get the correct uncertainties from it if samples composed of multiple subsamples with different relative weights are used. All simulated samples used in this analysis are combined samples, so the uncertainties on the fractions arising from the limited statistic precision of the simulated samples, have to be determined with an alternative method which is described below.

### Result of the fit

Only the fractions of the  $W$ +jets and top contributions are fitted. The fraction of the  $Z$ +jets and QCD multijet contributions are constrained and only allowed to fluctuate within their respective uncertainties. Figure 6.11 shows the result of the fit to the distribution in data compared with the sum of the initial Monte Carlo distribution before the fit. The fit result is denoted in the figure with a dashed line. It can be seen that the agreement between data and Monte Carlo improves significantly. The  $\chi^2/\text{ndf}$  is 9.0/8 and the linear correlation between the two fit parameters  $p_W$  and  $p_{\text{top}}$  is -30.6%

The fractions obtained from the fit can be translated to event yields summarized in table 6.5.

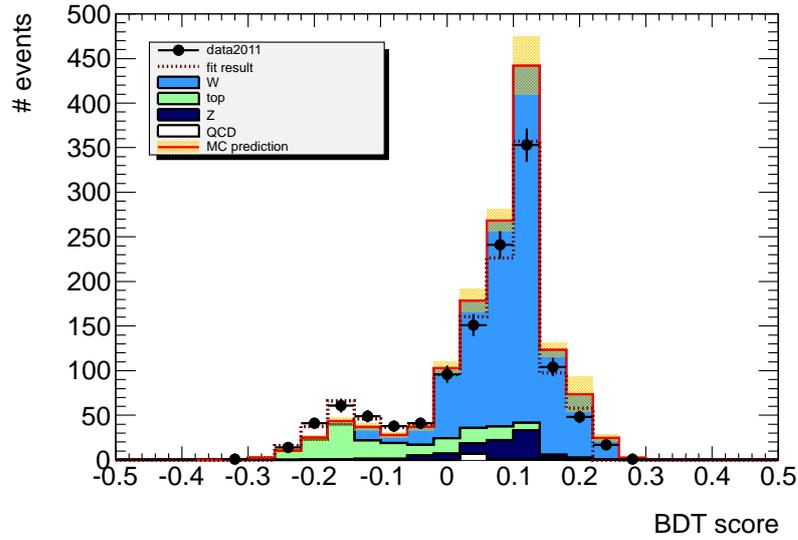


Figure 6.11: Output of the Barlow-Beeston fit to the BDT distribution in data in  $CR_1$ . The histograms show the unscaled contributions from the different Monte Carlo sources and the sum of these histograms is shown with a red line. The dotted line shows the sum of all Monte Carlo sources if the fractions obtained from the fit are applied.

As already mentioned, it is noticeable when comparing data with Monte Carlo in figure 6.10 that the data tends to overshoot in top-dominant regions. To make sure that this effect is not solely due to the difference between data and Monte Carlo in the number of selected  $b$ -jets, biasing the fit towards a higher top contribution, the training of the BDT discriminant was repeated without using the information from  $b$ -tagging and the fit performed with the resulting BDT.

The outcome of this fit can be seen in figure 6.12 and is also summarized in table 6.5. This result is consistent with the one obtained from the fit including the number of  $b$ -jets, but has a larger uncertainty, since the separation of the  $W$ +jets and top contributions is significantly worse when the  $b$ -tagging information is not used and the  $W$ +jets and top fit parameters are thus more highly correlated. The  $\chi^2/\text{ndf}$  of this fit is 5.67/8 with a linear correlation between the two fit parameters of -56.9%.

### Determination of the fit uncertainties

As mentioned above, the uncertainties on the fit parameters cannot be used out-of-the box because the distributions for each source are created from a combination of subsamples with different weights. Hence two methods for determining the uncertainties arising from

|           | Nominal fit  | Fit without $b$ -tagging | MC prediction |
|-----------|--------------|--------------------------|---------------|
| $W$ +jets | $849 \pm 48$ | $846 \pm 57$             | $1118 \pm 42$ |
| top       | $295 \pm 29$ | $297 \pm 45$             | $185 \pm 7$   |

Table 6.5: Results of two fits in  $CR_1$  compared to MC predictions. The two fits are consistent, but the fit without  $b$ -tagging information is less discriminating. The  $\chi^2/\text{ndf}$  is 9.0/8 (5.67/8) for the fit with (without)  $b$ -tagging information, and the linear correlation between the two fit parameters is -30.6% (-56.9%).

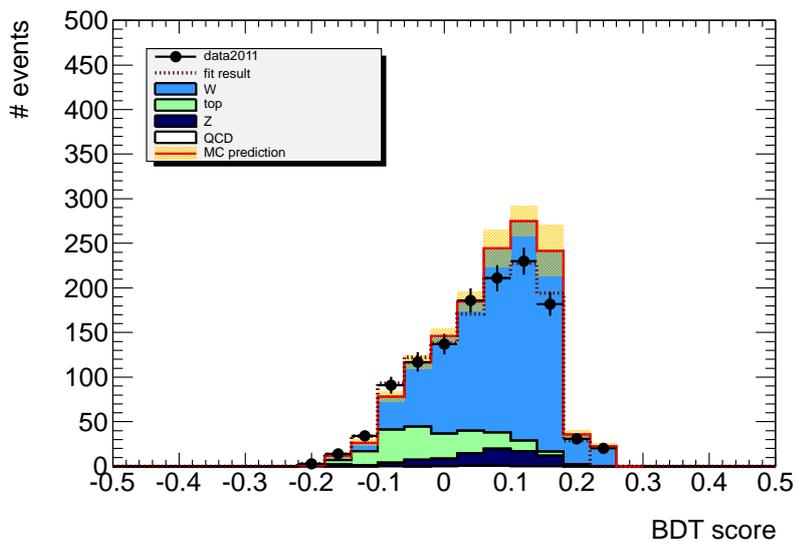


Figure 6.12: Output of the Barlow-Beeston fit to the BDT distribution in data, trained without the information of the number of  $b$ -jets. Again the dotted line denotes the distribution of the scaled sum of all Monte Carlo sources using the fractions from the Barlow-Beeston fit algorithm.

the statistical uncertainty of the data in  $CR_1$  and the from limited size of the Monte Carlo samples were investigated separately.

1. Uncertainty due to the number of data events  $\sigma_{data}$ :

To obtain an estimate for the uncertainty due to the number of data events only, the Monte Carlo distributions of the BDT discriminant are scaled by a factor in the range from 2 to 100. The fit is then performed with the changed distributions and it is determined when the Monte Carlo uncertainties become a negligible effect on the fit results. Figure 6.13 shows the evolution of the uncertainties evaluated from the fit for the two parameters. It is seen that the uncertainties do not change for a scale factor of  $\sim 30$  and above. This uncertainty is then taken as  $\sigma_{data}$ . The reason for not just scaling the samples by a very large number is that the fit gets numerically unstable for very large scaling factors and thus the calculated uncertainties cannot be trusted anymore.

2. Uncertainty due to limited size of simulated Monte Carlo samples  $\sigma_{MC}$ :

To estimate the uncertainty due to the limited statistical precision of the simulated Monte Carlo samples, a toy study has been performed. Instead of obtaining the fractions for the  $W$ +jets and top contributions from a fit to ATLAS data, a toy-data set is built by fluctuating all simulated input templates separately in each bin, according to a Poisson distribution. 10000 such toy-data distributions are created.

The fluctuated Monte Carlo sources are added up. The unfluctuated Monte Carlo templates are then used to fit the toy-data distribution and to obtain the fractions for the  $W$ +jets and top contributions.

Figure 6.14 shows the result the 10000 fits to toy-datasets. From the widths of this distributions, obtained by fitting a normal distribution to the resulting histogram, two uncertainties are extracted.

The two uncertainties  $\sigma_{data}$  and  $\sigma_{MC}$  are added in quadrature for  $W$ +jets and top separately. The resulting total uncertainties are given in table 6.5.

Using the nominal results, two scale factors are derived for the  $W$ +jets and top contributions in  $CR_1$  by comparing the yields obtained from the fit and the expectation from Monte Carlo:

$$f_i = \frac{\text{\# of fitted events in sample } i}{\text{\# of expected events from MC in sample } i} \quad (6.15)$$

The scale factors for  $W$ +jets and top are:

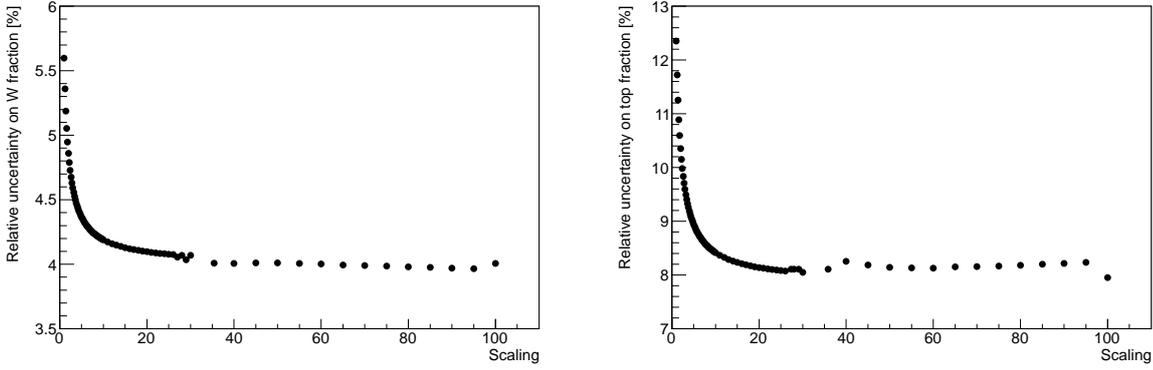


Figure 6.13: Evolution of the uncertainty on the two fit parameters.  $W$ +jets contribution (left) and top contribution (right) for different scalings. Both uncertainties reach a constant value for a scale factor of  $\sim 30$ . This value is taken as the uncertainty of the  $W$ /top fit parameter due to the statistical uncertainty of the data in  $CR_1$ .

$$f_W = 0.76 \pm 0.03 \quad (6.16)$$

$$f_{top} = 1.59 \pm 0.15. \quad (6.17)$$

which lead to the following estimate for the number of  $W$ +jets and top event yields with true tau leptons in the *signal region*:

$$N_{W,true}^{SR} = (4.0 \pm 1.4) \text{ events} , \quad (6.18)$$

$$N_{top,true}^{SR} = (3.7 \pm 1.5) \text{ events} . \quad (6.19)$$

The uncertainties quoted include the statistical uncertainties on the number of Monte Carlo events in the signal region and the uncertainties on the scale factors.

Figure 6.15 shows the input variables to the BDT and the BDT itself after applying the fitted scale factors to Monte Carlo in  $CR_1$ . A significantly improved agreement between data and Monte Carlo is seen in all of these distributions.

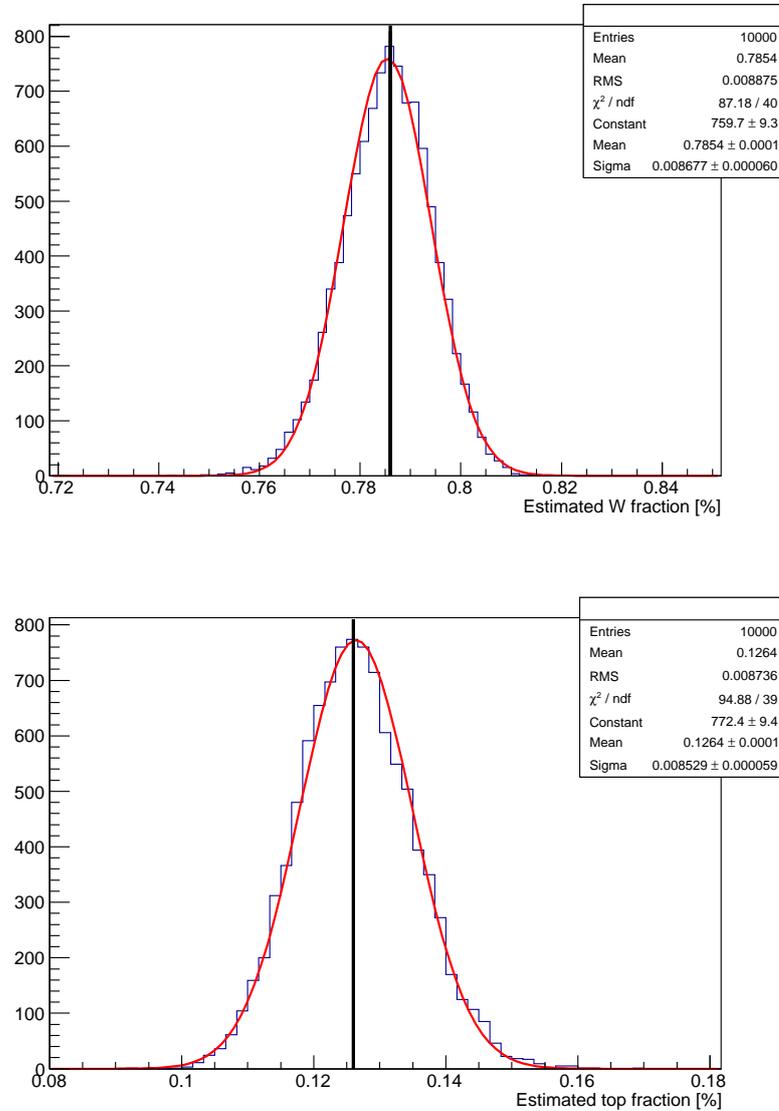


Figure 6.14: Result of 10000 fits of the input templates toy-data distributions for the  $W$ +jets fraction (left) and the top fraction (right). The black lines indicate the true fractions from Monte Carlo simulation. The uncertainties due to the limited number of simulated Monte Carlo events is taken from a Gaussian fit to the distributions. The fitted parameters of the normal distributions are also shown.

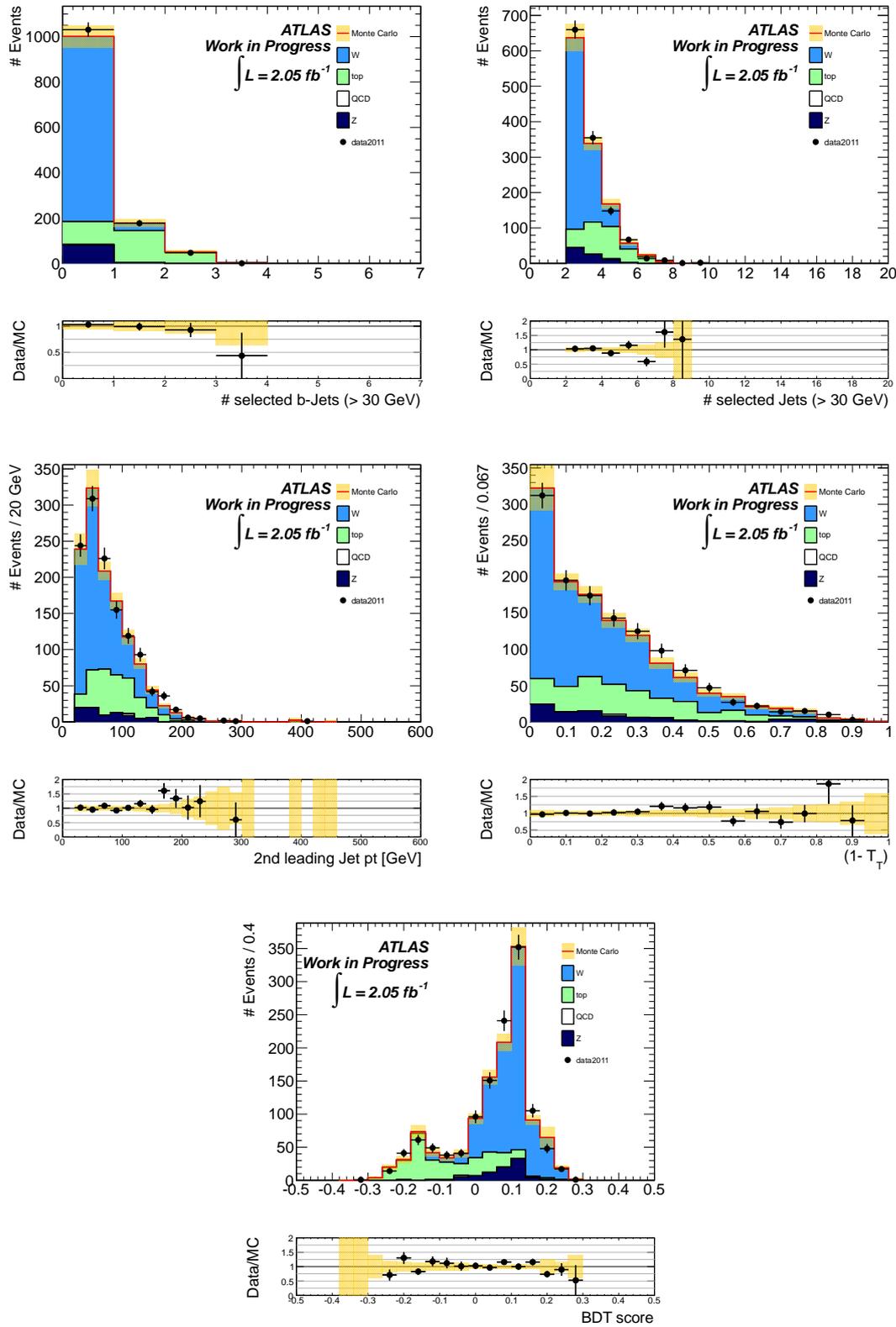


Figure 6.15: Input variables to the  $W$ /top separation BDT and the BDT discriminant itself, shown in the true-tau control region  $CR_1$  after applying the scale factors obtained from the Barlow-Beeston fit. The errorband shows the combined statistical uncertainty on the scale factors obtained from the fit.

|           | identified tau lepton | misidentified tau candidate | total            |
|-----------|-----------------------|-----------------------------|------------------|
| $W$ +jets | $102.7 \pm 7.8$       | $84.4 \pm 7.0$              | $187.1 \pm 10.5$ |
| top       | $58.9 \pm 7.8$        | $62.1 \pm 8.2$              | $121.0 \pm 11.3$ |
| $Z$ +jets | $5.4 \pm 1.7$         | $70.0 \pm 13.5$             | $75.4 \pm 13.6$  |
| QCD       | - $\pm$ -             | $6.6 \pm 2.7$               | $6.6 \pm 2.7$    |
| total     | $167.0 \pm 11.2$      | $223.1 \pm 17.5$            | $390.1 \pm 20.7$ |
| data      |                       |                             | 350              |

Table 6.6: Numbers of observed and expected events in  $CR_2$ . The numbers of expected  $W$ +jets and top events have been corrected by the scale factors measured in  $CR_1$ .

## 6.2 $W$ +jets / top background with misidentified tau candidates

The low- $m_T$  ( $\tau, E_T^{\text{miss}}$ ) control region  $CR_1$  is suitable to obtain scale factors for  $W$ +jets and top events with identified tau leptons. To estimate the background with misidentified tau candidates a different control region,  $CR_2$ , (see figure 6.2) is used. From figure 6.1 it is obvious that only a small fraction of  $W$ +jets and top events with truth-matched tau candidates does not end up in  $CR_1$ . Table 6.6 shows the number of events in  $CR_2$  split up in contributions from identified tau leptons and misidentified tau candidates. The numbers of events from  $W$ +jets and top are already scaled by the factors derived in  $CR_1$ .

It can be seen that, although the number of events with misidentified tau candidate is substantially larger than in  $CR_1$ , there are also sizeable contributions from events with truth-matched tau candidates. In order to calculate a scale factor for events with misidentified tau candidates only, the expected contribution from events with truth-matched tau candidates is subtracted from the data and compared with the total expectation of Monte Carlo events with misidentified tau candidates.

Unlike the procedure done in  $CR_1$  this scale factor characterizes all non-QCD events with misidentified tau candidates instead of using different scale factors for different physics processes.

$$f_{\text{fake}} = \frac{N^{\text{CR}_2} - \hat{N}_{\text{true}}^{\text{CR}_2}}{\hat{N}_{\text{fake}}^{\text{CR}_2}} \quad (6.20)$$

Using table 6.6 and equation 6.20, the fake-tau scale factor  $f_{\text{fake}}$  is obtained:

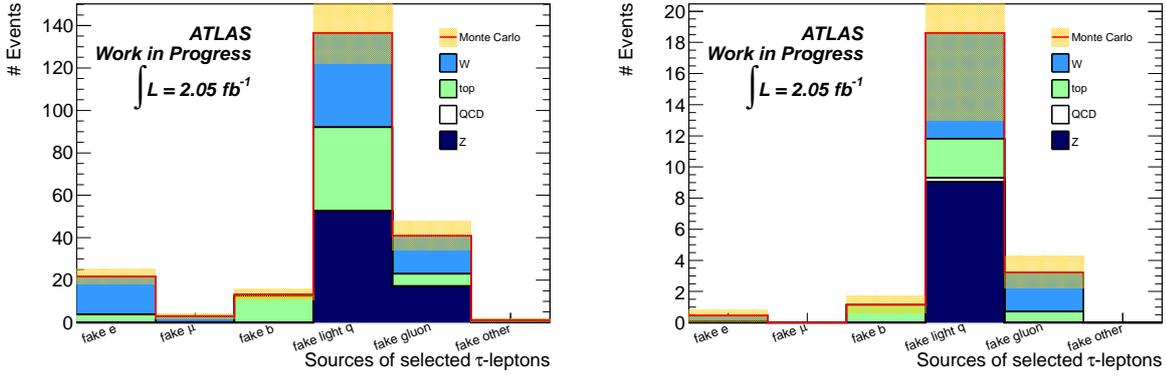


Figure 6.16: Comparison of the sources of misidentified tau candidates between CR<sub>2</sub> (left) and the signal region (right). The different categories from left to right are: tau candidates from electrons, tau candidates from muons, tau candidates from  $b$ -quarks, tau candidates from light quarks, tau candidates from gluons and unmatched tau candidates.

$$f_{fake} = \frac{350 - (167.0 \pm 11.2)}{223.1 \pm 17.5} = 0.82 \pm 0.10 \quad (6.21)$$

The uncertainties on the numbers are the combined uncertainty due to the number of Monte Carlo and data events and the uncertainties on the  $W$ +jets and top scale factors derived in CR<sub>1</sub>.

To use this scale factor to correct the background from misidentified tau candidates in the signal region, it needs to be checked that the misidentified tau candidates described in CR<sub>2</sub> are similar to the ones in the signal region with respect to the parameters related to the misidentification probability of the tau identification.

As described in section 4.2, the tau identification is particularly sensitive to the sources and the transverse momentum of the misidentified tau candidates. Figures 6.16 and 6.17 show the source split up in different categories and the transverse momentum of the misidentified tau candidate in CR<sub>2</sub> and in the signal region side by side.

It is found that the two regions in phase space are quite similar with respect to the sources and the transverse momentum of the misidentified tau candidates which gives confidence that the scale factor for misidentified tau candidates derived in CR<sub>2</sub> can be used to model the contribution from misidentified tau candidates in the signal region.

When applying this scale factor to the Monte Carlo expectation of events with misidentified tau candidates from  $W$ +jets and top in the signal region an event yield of

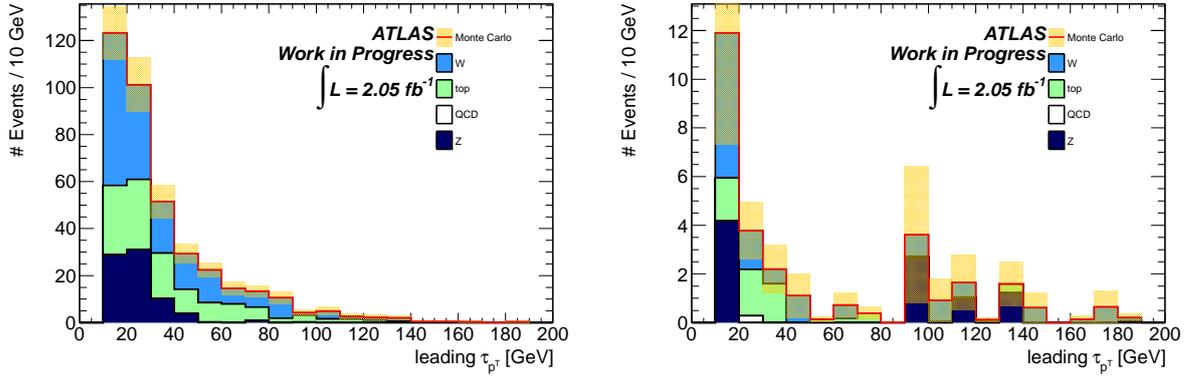


Figure 6.17: Comparison of the  $p_T$  of the misidentified tau candidates between CR<sub>2</sub> (left) and the signal region (right).

$$\hat{N}_{W + \text{top, fake}}^{\text{SR}} = (11.8 \pm 2.5) \text{ events} \quad (6.22)$$

is obtained. The uncertainty includes the statistical uncertainty on the Monte Carlo samples, the uncertainties on the three scale factors,  $f_W$ ,  $f_{\text{top}}$  and  $f_{W + \text{top, fake}}$  and an additional systematic uncertainty on the mixture of the various backgrounds in the Monte Carlo. The latter uncertainty is estimated by varying the relative fractions of  $W$ +jets, top and  $Z$ +jets events with misidentified tau candidates by 50% up and down and the relative fraction of QCD-multijet events by 100%. Then the sum of background events with misidentified tau candidates in the signal region is recalculated. On average this adds an additional uncertainty of  $\sim 6\%$ .

# Chapter 7 Estimation of the QCD-multijet background

---

This chapter presents a data-driven technique for estimating the number of background events from QCD-multijet production.

Due to the large production cross-section and the small number of Monte Carlo events after requiring an identified tau lepton, a dedicated data-driven estimation is mandatory to sufficiently understand the background contribution from QCD-multijet events in the signal region. The presented method gives only a rough estimate, which is sufficient, because the QCD-multijet background in the signal region is expected to be small compared to the other backgrounds in the analysis.

A first attempt to obtain an estimate of the QCD-multijet background in the signal region was presented in section 5.3.1. For that method, a separate dijet selection was used to determine the tau lepton misidentification probabilities from data and weigh the Monte Carlo according to these. To determine the QCD-multijet background from data a so-called *matrix-method* is presented. First an overview of the concept of the matrix method is explained in section 7.1 and some modifications to the method that need to be applied are discussed in section 7.2. At the end of the chapter, studies to validate the method are presented.

The method of reweighting the Monte Carlo according to the misidentification probabilities is needed mainly to have a large enough Monte Carlo sample to investigate the requirements to reject the QCD-multijet background in the analysis. Furthermore it is needed as a starting point for the estimation of the  $W$ +jets and top scale factors, since they are used for this data-driven estimation of the QCD-multijet background. Of course they can also be used to cross-check the results obtained with this method. It is also important to have the dedicated QCD-estimation in  $CR_1$ , since the estimation presented here can not be used in the low  $m_T(\tau, E_T^{\text{miss}})$  region.

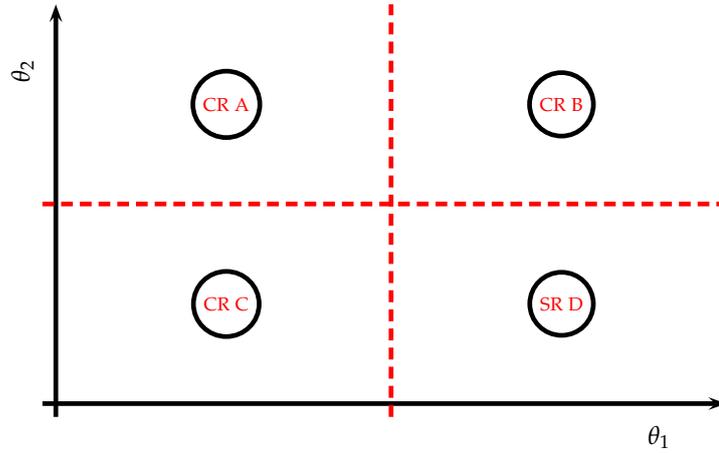


Figure 7.1: Schematic view of the control regions (CR) A–C and the signal region (SR) D in the  $\theta_1$ - $\theta_2$  parameter space.

## 7.1 Overview of the method

The general strategy of the data-driven QCD-multijet estimation is to define a QCD enhanced region in phase space that includes the signal region and can be subdivided into three QCD-enhanced control regions, **A**, **B** and **C** and the signal region, **D**, by requirements on two uncorrelated variables  $\theta_1$  and  $\theta_2$ . This is shown in a schematic view in figure 7.1.

The number of data events from one of the regions separated by only one requirement (either on  $\theta_1$  or  $\theta_2$ ) from the signal region **D**, e.g. region **C** is extrapolated to the signal region by multiplying the number of QCD-multijet events in control region **C** by the ratio of QCD-multijet events in regions **B** and **A**. The same can be done by using the number of QCD-multijet events in region **B** and multiplying it by the ratio of QCD-multijet events in **A** and **C**.

Two assumptions are made for this procedure:

- The ratio of the number of QCD-multijet events in region **B** and **A** is only equal to the ratio of number of QCD-multijet events in regions **D** and **C** if the two variables  $\theta_1$  and  $\theta_2$  are uncorrelated.
- If shapes are plotted of variables for the estimated QCD-multijet background the shape of the variable in region **C** has to be equal to the shape of the variable in the signal region **D**.

To define a region in phase space with both a large number and high purity of QCD-multijet events, the tau identification is loosened. This is done by applying a very loose preselection on all tau candidates in the event. The preselection criteria are:

- $|q|_{\text{tau candidate}} < 3$  and
- $N_{\text{tracks, tau candidate}} > 6$ .

All tau candidates passing these preselection criteria are in the following called *loose* tau candidates.

To be able to suppress non-QCD backgrounds, one of these loose tau candidates is chosen randomly in every event. The event is then propagated through the selection, omitting the requirements on the angular separation between the first two leading jets and  $E_T^{\text{miss}}$  in the event,  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}})$ , because they are later used as one of the variables to split the sample into the four regions.

The randomly chosen tau candidate is needed to be able to apply the requirement on  $m_T(\tau, E_T^{\text{miss}})$  to reduce the  $W$ +jets and top backgrounds.

The omitted requirements on  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}})$  and the identification decision for the randomly selected tau candidate are then used to divide the obtained sample into four regions: **A**, **B**, **C** and **D**.

- Regions **A** and **C** contain events where the randomly chosen tau candidate **fails the tau identification criteria**.  
Events where the tau candidate **passes the tau identification** are included in regions **B** and **D**.
- Events which **pass the requirements** on  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}}) > 0.3$  rad are included in regions **A** or **B**.  
Regions **C** and **D** contain events which **fail the requirement** on  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}})$ .

This makes region **D** the *signal region*, corresponding to the nominal selection. Regions **A**, **B** and **C** are *control regions*. The definition of the different regions is summarized in table 7.1.

Table 7.2 shows the number of data and simulated events for the four different regions

To obtain an estimate for the number of QCD-multijet events in the signal region, the number of QCD events in region **C** is transferred to the signal region **D** by multiplying with the ratio  $N_{\text{QCD}}^{\text{B}}/N_{\text{QCD}}^{\text{A}}$ :

$$\hat{N}_{\text{QCD}}^{\text{D}} = N_{\text{QCD}}^{\text{C}} \times \frac{N_{\text{QCD}}^{\text{B}}}{N_{\text{QCD}}^{\text{A}}} . \quad (7.1)$$

This equation only holds if there are no correlations between the variables used to split the samples. Also, only the number of the sum of all contributions in the different regions is known. To obtain the number of QCD-multijet events in the three control regions, the

|                                    | fail tau identification<br>$N_{trk} < 6,  q  < 3$ | pass tau identification<br>$N_{trk} = 1 \text{ or } 3,  q  = 1$ |
|------------------------------------|---|---|
| Small $\Delta\phi$ (inverted cuts) | Control region A                                  | Control region B  |
| Large $\Delta\phi$ (nominal cuts)  | Control region C                                  | Signal region D   |

Table 7.1: Definition of signal and control regions for the QCD-multijet background estimation method.  $\Delta\phi$  stands symbolically for the requirements on the angular separation between  $E_T^{\text{miss}}$  and the first two leading jets in the event. The requirements on  $N_{\text{track}}$  and  $|q|$  are not additional requirements but are used to identify tau candidates in the nominal selection (see section 4.2).

|          | data | W+jets   | top      | Z+jets   | QCD       |
|----------|------|----------|----------|----------|-----------|
| Region A | 4643 | 273±34   | 204±29   | 230±37   | 4109±697  |
| Region B | 22   | 2.5±1.4  | 2.7±1.0  | 6.3±5.1  | 0.13±0.06 |
| Region C | 6612 | 1980±226 | 1359±187 | 2087±232 | 517±257   |
| Region D | 57   | 21.1±4.7 | 21.9±3.7 | 20.3±7.9 | 0.28±0.28 |

Table 7.2: Numbers of data and simulated events in the four regions of the QCD background estimation method. The Numbers for W+jets and top are scaled with the factors obtained in chapter 6. The errors shown include Monte Carlo statistical uncertainties and statistical uncertainties on the scale factors.

contributions from non-QCD backgrounds need to be subtracted.

## 7.2 Modification of the method

In order to account for the relatively small number of events in control region **B** and the contamination from non-QCD backgrounds in the control regions, modifications to the matrix-method have to be introduced:

- Small number of events in control region **B**:

Unfortunately, with the dataset used for the analysis, the number of events in the control regions, especially in region **B**, is small. To increase the number of QCD-multijet events, the requirement on  $E_T^{\text{miss}}/m_{\text{eff}}$  is loosened in all regions with respect to the nominal analysis from  $E_T^{\text{miss}}/m_{\text{eff}} > 0.25$  to  $E_T^{\text{miss}}/m_{\text{eff}} > 0.1$ .

- Contamination from  $W$ +jets, top and  $Z$ +jets events in the control regions:

While regions **A** and **B** are clearly dominated by QCD-multijet events after the loosened requirement on  $E_T^{\text{miss}}/m_{\text{eff}}$ , there is a sizeable contamination from  $W$ +jets, top and  $Z \rightarrow \nu\nu$ +jets in region **C**. To obtain the number of QCD-multijet events in the three control regions, the number of non-QCD background events is subtracted from data in each region separately.

Figure 7.2 shows the distribution of  $E_T^{\text{miss}}/m_{\text{eff}}$  in region **C**. It can be seen that a large amount of non-QCD background has to be subtracted from data, which results in a large uncertainty on the number of QCD-multijet events.

In order to reduce the amount of contamination that needs to be subtracted, an additional requirement on  $E_T^{\text{miss}}/m_{\text{eff}} < 0.25$ , which is the reversed requirement with respect to the nominal analysis used there to reject QCD-multijet background, is set only to events in region **C**, indicated in figure 7.2 by the black line with an arrow. This requirement reject a large fraction of the contamination from non-QCD backgrounds.

For the final estimate of the number of QCD-multijet events in signal region **D**, an extrapolation from the restricted  $E_T^{\text{miss}}/m_{\text{eff}}$  region ( $0.1 < E_T^{\text{miss}}/m_{\text{eff}} < 0.25$ ) to the one used in the nominal selection ( $E_T^{\text{miss}}/m_{\text{eff}} > 0.25$ ) needs to be performed.

The numbers of data and simulated events split up in the four different regions with the additional requirements are summarized in table 7.3. The Monte Carlo expectations for  $W$ +jets, top and  $Z$ +jets production are scaled using the factors derived in chapter 6. Figure 7.3 shows the  $m_{\text{eff}}$  distribution in the four different regions. Clear normalization

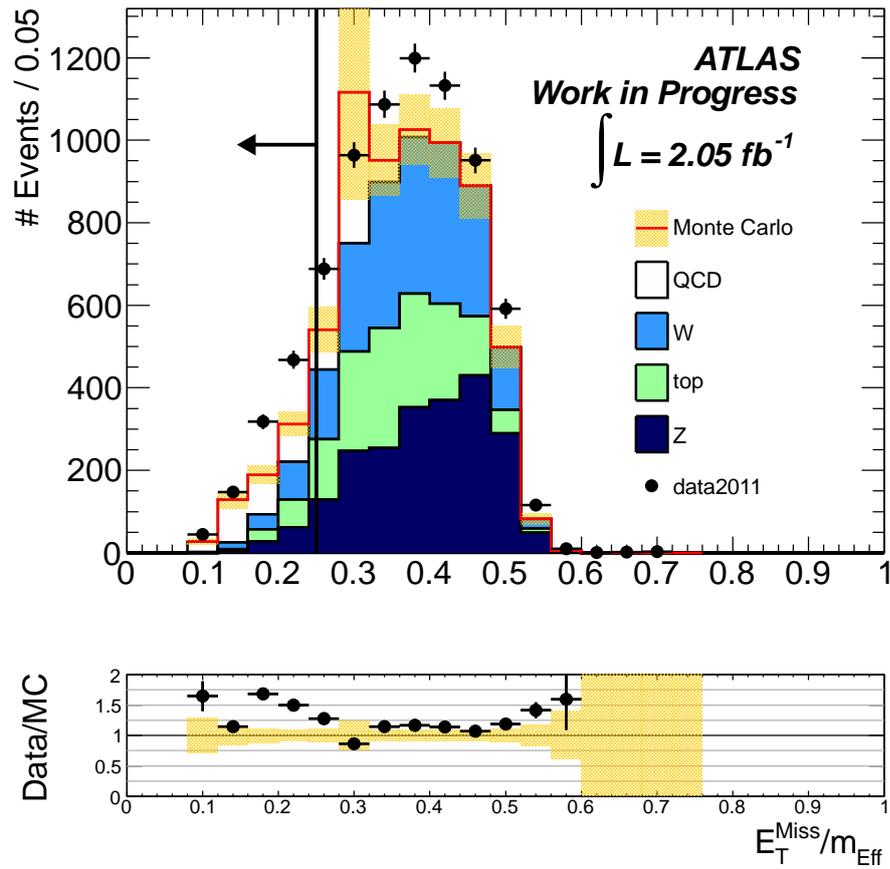


Figure 7.2: Distribution of  $E_T^{\text{miss}}/m_{\text{eff}}$  in control region C. The additional requirement intended to reduce the contamination from non-QCD backgrounds - especially  $Z \rightarrow \nu\nu$  - is indicated by the vertical line.

|          | data  | W+jets     | top        | Z+jets     | QCD        |
|----------|-------|------------|------------|------------|------------|
| Region A | 11085 | 344.9±41.6 | 266.3±42.5 | 272.6±40.7 | 9234±861   |
| Region B | 42    | 3.0±1.6    | 3.6±1.2    | 6.3±5.0    | 110±90     |
| Region C | 1111  | 185.5±51.1 | 122.8±30.0 | 105.1±45.2 | 333.7±35.3 |
| Region D | 60    | 22.5±4.8   | 21.51±4.0  | 22.7±8.4   | 4.5±4.1    |

Table 7.3: Numbers of data and simulated events in the four regions of the QCD background estimation method. In contrast to table 7.2, the selection requirements are changed as discussed in section 7.2.

differences between data and Monte Carlo, but also some differences in the shapes of the distributions at the 5% to 30% level is seen.

The numbers of QCD events in regions **A–C** are obtained by subtracting from data the contamination due to non-QCD backgrounds using the Monte Carlo prediction scaled with the factors derived in chapter 6. These “corrected” numbers are then used to obtain  $N_{\text{QCD}}^D$  in the intermediate  $E_T^{\text{miss}}/m_{\text{eff}}$  region:

$$\begin{aligned}
\hat{N}_{\text{QCD}}^D \Big|_{0.1 < E_T^{\text{miss}}/m_{\text{eff}} < 0.25} &= \frac{N_{\text{corrected}}^B}{N_{\text{corrected}}^A} \Big|_{0.1 < E_T^{\text{miss}}/m_{\text{eff}}} \times N_{\text{corrected}}^C \Big|_{0.1 < E_T^{\text{miss}}/m_{\text{eff}} < 0.25} \\
&= \frac{29.1}{10201.1} \times 697.6 \\
&= (2.0 \pm 0.6) \text{ events} .
\end{aligned} \tag{7.2}$$

The uncertainty on this value includes the statistical uncertainties on the data in the different regions, the uncertainty due to the limited size of the Monte Carlo samples and the uncertainties on the factors used to scale the non-QCD samples.

An additional uncertainty on the number of W+jets, top and Z+jets in region **C** is included to account for a possible mismodelling of the tau identification in the simulation. Since the requirement on the tau identification is used to separate region **C** and region **D**, 100% of the contribution in the signal region **D**, corresponding to additional 22.5 events for W+jets, 21.5 events for top and 22.7 events for Z+jets, is added to the total uncertainty of these samples in region **C**. This translates to an uncertainty on the tau misidentification probability of  $\sim 16\%$ .

The result in the intermediate  $E_T^{\text{miss}}/m_{\text{eff}}$  region needs to be transferred to the high  $E_T^{\text{miss}}/m_{\text{eff}} > 0.25$  region. The transfer factors needed to calculate the number of QCD-multijet after the

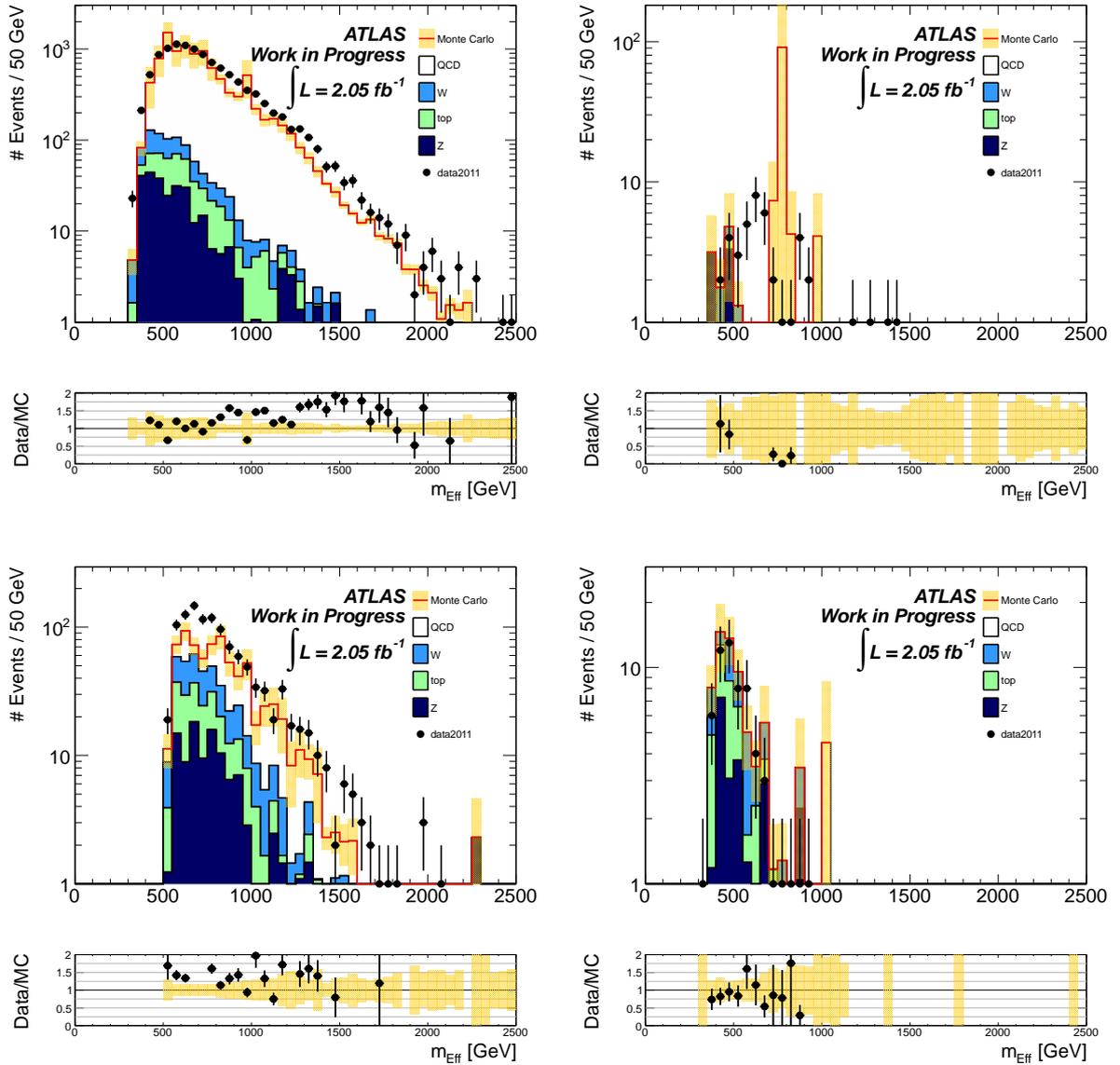


Figure 7.3: Distribution of the effective mass in the four different regions of the QCD-multijet background estimation method. Upper left: control region **A**, upper right: control region **B**, lower left: control region **C**, lower right: signal region **D**. The non-QCD backgrounds are scaled using the factors obtained in chapter 6. The errors shown include both the Monte Carlo statistical uncertainties and the uncertainties on the scale factors.

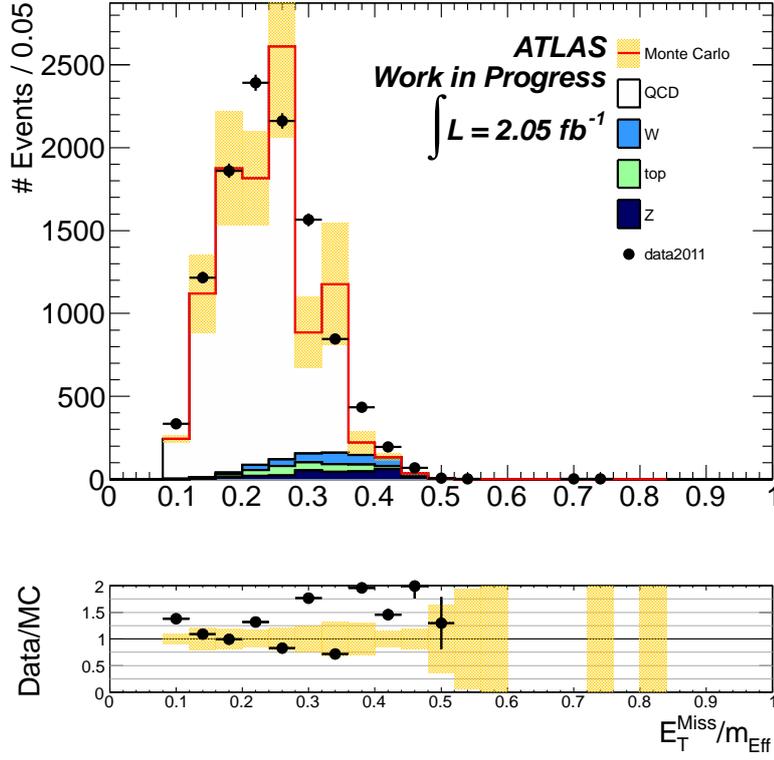


Figure 7.4:  $E_T^{\text{miss}}/m_{\text{eff}}$  distribution in control region A. This distribution is used to calculate the extrapolation factor from the intermediate  $E_T^{\text{miss}}/m_{\text{eff}}$  region to the signal region.

$m_T(\tau, E_T^{\text{miss}})$  requirement and the final requirement  $m_{\text{eff}} > 600 \text{ GeV}$ , is obtained in control region A and amount to  $0.63 \pm 0.02$  and  $0.200 \pm 0.010$  respectively. The  $E_T^{\text{miss}}/m_{\text{eff}}$  distribution used to calculate the extrapolation factors is shown in figure 7.4.

The resulting QCD background estimates are

$$\hat{N}_{\text{QCD}}^D \Big|_{E_T^{\text{miss}}/m_{\text{eff}} > 0.25} = (1.25 \pm 0.37) \text{ events} \quad (7.3)$$

for the number of events after the requirement on  $m_T(\tau, E_T^{\text{miss}})$ , and

$$\hat{N}_{\text{QCD}}^{\text{SR}} = \hat{N}_{\text{QCD}}^D \Big|_{E_T^{\text{miss}}/m_{\text{eff}} > 0.25 \& m_{\text{eff}} > 600 \text{ GeV}} = (0.39 \pm 0.12) \text{ events} \quad (7.4)$$

in the signal region. This result is consistent with the one obtained by the Monte Carlo weighting method (see table 5.4) which yields  $(0.4 \pm 0.1)$  events in the signal region.

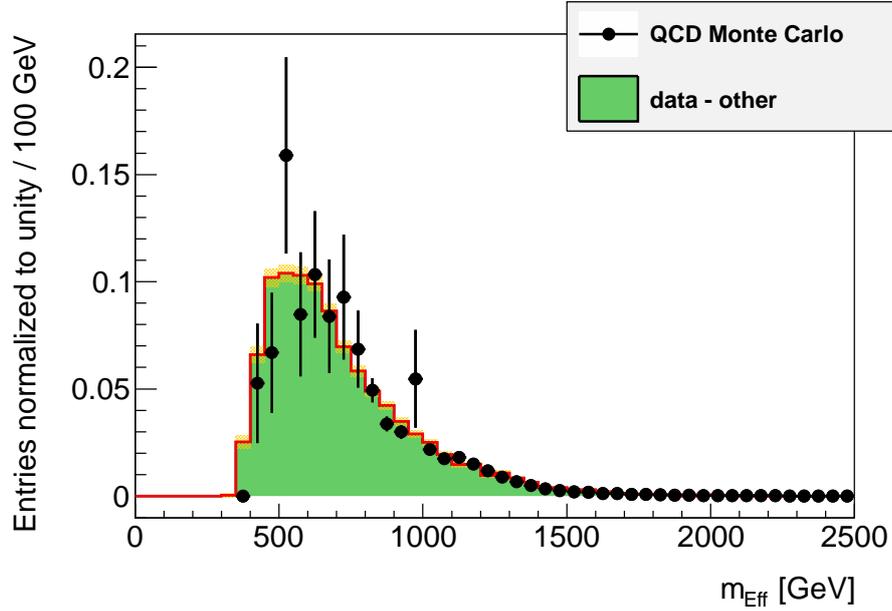


Figure 7.5: Comparison of the  $m_{\text{eff}}$  distributions for QCD Monte Carlo and the QCD estimate from data, with the non-QCD backgrounds subtracted off. The histogram is shown for all four regions (A,B,C,D) combined in order to have a reasonable statistical precision. Note: The Monte Carlo expectation is shown as black closed circles with error bars.

### 7.3 Validation of the method

To ensure that all assumptions made in the QCD estimation method are fulfilled and to test the statistical robustness of the method several validation studies have been done.

As a first test, the sample obtained by subtracting the non-QCD backgrounds from data is compared with the QCD Monte Carlo. Figure 7.5 shows the  $m_{\text{eff}}$  distribution for this sample together with the distribution from QCD-multijet Monte Carlo simulation for the four regions of the matrix method combined. A good agreement between the two distributions is seen within the large statistical uncertainties of the QCD simulation.

#### 7.3.1 Check for correlations

Since the matrix method requires that the variables used to split the sample in the four regions are uncorrelated, the correlation between  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}})$  and the tau identification  $\text{BDT}_{\text{tau}}$  has been studied. Figure 7.6 shows the distribution of  $\text{BDT}_{\text{tau}}$  in slices of  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}})$  for the  $W$ +jets and QCD-multijet Monte Carlo samples. Three regions of  $\Delta\phi$  are chosen to have approximately equal number of events. The chosen regions are

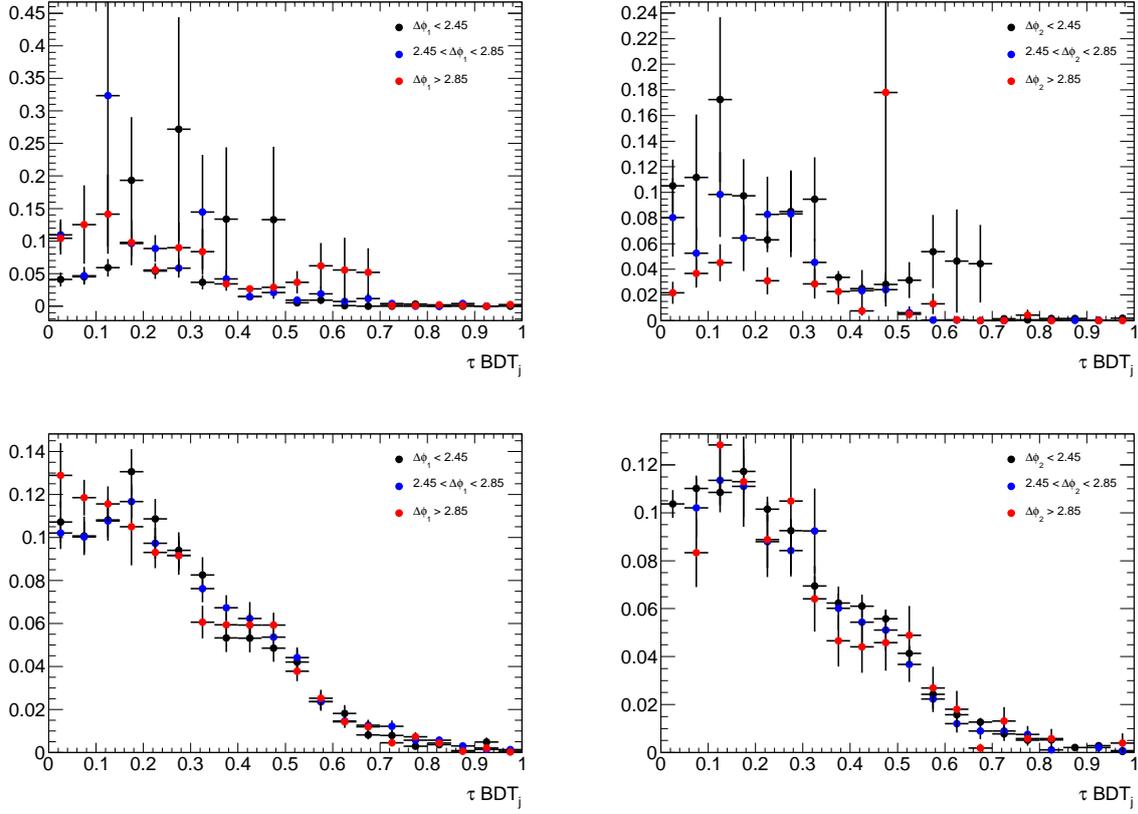


Figure 7.6: Distributions of the tau identification  $BDT_{\tau_{au}}$  in slices of  $\Delta\phi$ . Three regions of  $\Delta\phi$  are chosen to have approximately equal numbers of events. They are shown in different colors. The plots on the left show  $\Delta\phi$  for the leading jet, while those on the right show  $\Delta\phi$  for the second-leading jet. The top plots show the QCD-multijet Monte Carlo while the bottom shows the  $W$ +jets simulated sample.

$\Delta\phi < 2.45 \text{ rad}$ ,  $2.45 \text{ rad} < \Delta\phi < 2.85 \text{ rad}$ ,  $\Delta\phi > 2.85 \text{ rad}$ . No significant correlations are seen for QCD-multijet and  $W$ +jets simulation, but the statistical uncertainties are still sizeable. The same has been found for the backgrounds from top and  $Z$ +jets.

As a second test of the correlations, the  $m_{\text{eff}}$ -distributions are compared between the combined regions **A+B** and **C+D** and between regions **A+C** and **B+D**. The results are shown in figure 7.7 for data with the contributions from  $W$ +jets, top and  $Z$ +jets subtracted off. It can be seen that for both figures the two distributions are somewhat shifted with respect to each other. Since the effective mass is calculated with the  $p_T$  of the two leading jets in the event, the shift in the distributions of the combined regions **A+B** ( $E_T^{\text{miss}}$  due to mismeasurement of one of the two leading jets) and **C+D** (no angular correlation between the first two leading jets and  $E_T^{\text{miss}}$ ) can be understood. If  $E_T^{\text{miss}}$  can be regarded as a mismeasurement

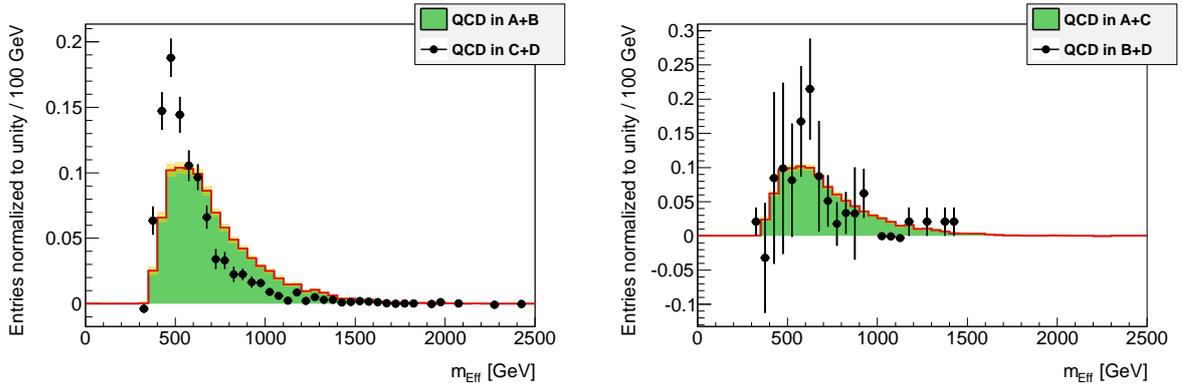


Figure 7.7: Comparison of the shapes between the  $m_{\text{eff}}$ -distributions in the combined regions with low  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}})$  **A+B** and high  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}})$  **C+D** (left) and the combined regions with failed tau identification **A+B** and passed tau identification **C+D** (right)

of one of the two leading jets this jet needs to have large initial  $p_T$ . Since the jets in QCD-multijet events need to be balanced in  $p_T$ , the second jet also tends to have very large  $p_T$ . This shifts the  $m_{\text{eff}}$  distribution to larger values.

The difference in the distributions of the combined region **A+C** and **B+D** can be explained with the strong dependency of the tau misidentification probability on the  $p_T$  of the tau candidate which is also used to calculate the effective mass.

These discrepancies cause a problem, because the transfer factor from the intermediate  $E_T^{\text{miss}}/m_{\text{eff}}$  region to the high  $E_T^{\text{miss}}/m_{\text{eff}}$  region is taken from the control region **A**. To account for this, the transfer factor is also calculated in control region **B**, which has the same identification criteria applied to the loose tau candidate as in the signal region **D**. The transfer factors from region **B** amount to  $0.6 \pm 1.0$  for the QCD-multijet background after the  $m_T(\tau, E_T^{\text{miss}})$  requirement and  $0.5 \pm 0.4$  for the signal region. Figure 7.8 shows the  $E_T^{\text{miss}}/m_{\text{eff}}$  distribution in control region **B** used to calculate the transfer factors.

The number of expected QCD-multijet events in the signal region is recalculated using the transfer factor derived in region **B**. This yields an estimate of  $(1.1 \pm 0.9)$  events for the QCD-multijet background in the signal region. The very large uncertainty on this number is due to the small number of data events in region **B**. To account for the discrepancy in the two transfer factors the difference between the estimate in equation 7.4 and the one obtained with the transfer factor from control region **B** is taken as an additional uncertainty.

This results in the final estimate for the number of QCD-multijet events in the signal region of

$$\hat{N}_{\text{QCD}}^{\text{SR}} = (0.4 \pm 0.7) \text{ events} , \quad (7.5)$$

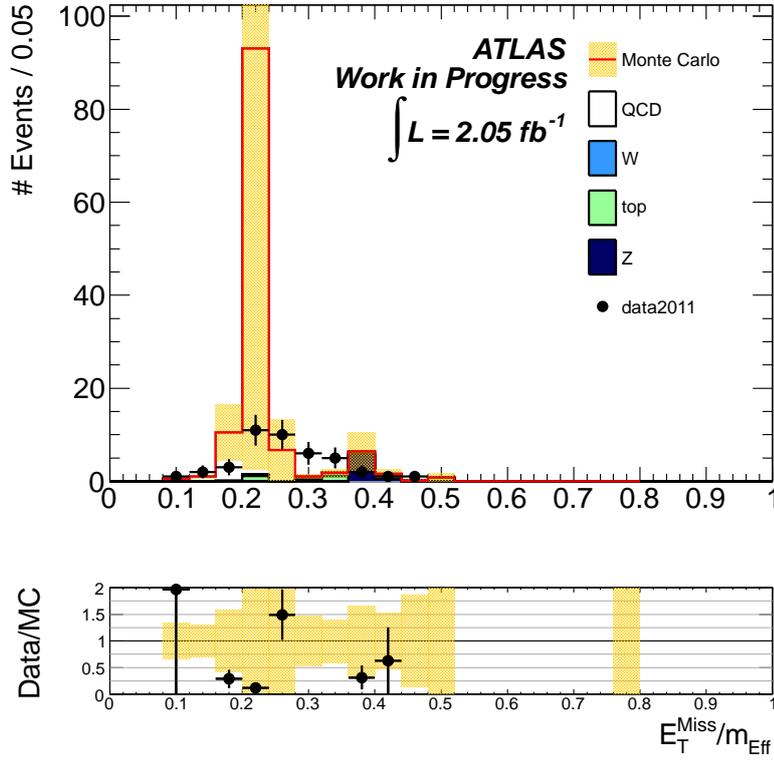


Figure 7.8:  $E_T^{\text{miss}}/m_{\text{eff}}$  distribution in control region B. This distribution is used to calculate the alternative transfer factor from the intermediate  $E_T^{\text{miss}}/m_{\text{eff}}$  region to the signal region.

where only the uncertainty has changed compared to the result from equation 7.4.

### 7.3.2 Test of the robustness

To validate the statistical robustness of the presented method a toy Monte Carlo study has been performed. The test data used for this study is obtained by fluctuating the sum of the non-QCD backgrounds within its uncertainty, as well as the QCD-multijet background obtained by subtracting off the non-QCD contributions from data and adding these two contributions up.

This is done in regions A–C independently. The number of events for the test data is then used to calculate the estimate  $\hat{N}_{\text{QCD}}^{D,fluc}$  by applying the full method described above. This is repeated 10000 times.

Figure 7.9 shows the distributions of the estimated value, the corresponding uncertainty of  $\hat{N}_{\text{QCD}}^{D,fluc}$  and the pull distribution defined by

$$w = \frac{\hat{N}_{fluc}^D - \hat{N}_{true}^D}{\sigma_{\hat{N}_{true}^D}}, \quad (7.6)$$

with  $\hat{N}_{true}^D$  being the result obtained with the nominal (unfluctuated) inputs and its uncertainty  $\sigma_{\hat{N}_{true}^D}$ .

For the number of estimated QCD-multijet events and its associated uncertainty the nominal value of  $(0.39 \pm 0.12)$  events from equation 7.4 is indicated by the black lines in the distributions. They both are located near the peak of the distributions.

The pull distribution is slightly asymmetric, but can still be reasonably well approximated by a normal distribution centered close to zero, indicating that the method is largely unbiased.

An additional study to evaluate the robustness of the method to changes of the requirements on  $E_T^{\text{miss}}/m_{\text{eff}}$  and  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}})$ , has been performed. For this study the values of the requirements on  $E_T^{\text{miss}}/m_{\text{eff}}$ ,  $\Delta\phi(\text{jet}_1, E_T^{\text{miss}})$  and  $\Delta\phi(\text{jet}_2, E_T^{\text{miss}})$  are changed independently. The lower requirement on  $E_T^{\text{miss}}/m_{\text{eff}}$  is varied in 5 steps from  $E_T^{\text{miss}}/m_{\text{eff}} > 0$  to  $E_T^{\text{miss}}/m_{\text{eff}} > 0.2$ . The requirements on the angular separation between the two leading jets and  $E_T^{\text{miss}}$  is changed from the nominal value of  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}}) > 0.3$  by a value of 0.05 up and down. In total, 45 different sets of requirements have been tested.

With these changed samples, the QCD-multijet estimation procedure is repeated and the results obtained are compared with the result of the nominal selection. Figure 7.10 shows the estimated number of QCD-multijet events and the uncertainty for the 45 variations. While there are some differences that can easily be understood since the definition of the signal regions changes (the signal region gets larger if the requirements on  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}})$  is loosened and smaller if it is tightened), it can be seen that the results are stable against the variations on these variables.

These studies show that the method yields a robust estimate for the QCD-multijet background.

## 7.4 Results

It was shown that with the matrix-method presented in this chapter, a data-driven estimate of the number of QCD-multijet background events in the signal region can be obtained. The estimate is largely unbiased and robust against statistical fluctuations and variations on the requirements on  $\Delta\phi(\text{jet}_{1,2}, E_T^{\text{miss}})$  and  $E_T^{\text{miss}}/m_{\text{eff}}$ . The final estimate of the number of QCD-multijet events in the signal region is:

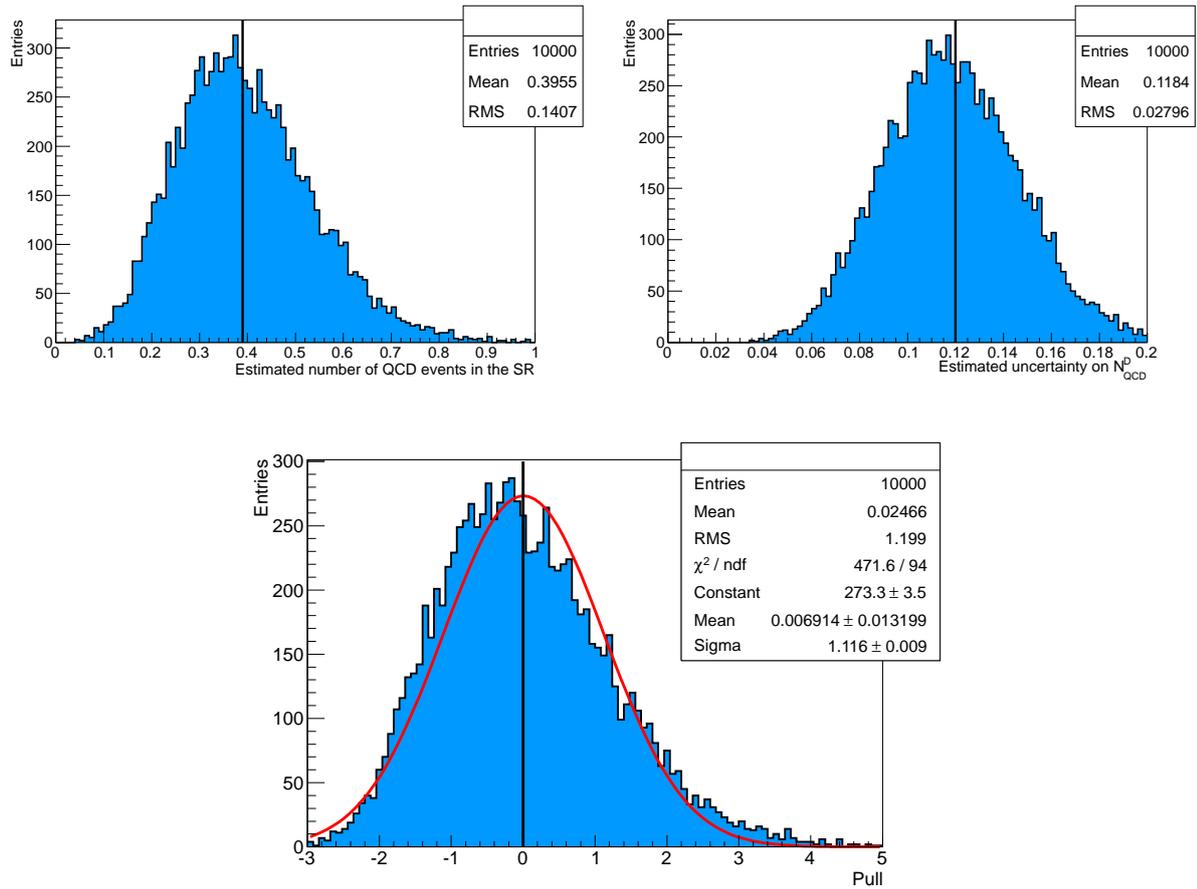


Figure 7.9: Results of 10000 toy MC experiments validating the matrix method for QCD-multijet background estimation. The number of expected QCD backgrounds extrapolated into the signal region (top left), its associated uncertainty (top right) and the pull distribution (bottom) are shown. The nominal values for the estimated number of QCD-multijet events and its uncertainty are indicated by the black lines. To guide the eye, the pull distribution is fitted to a normal distribution.

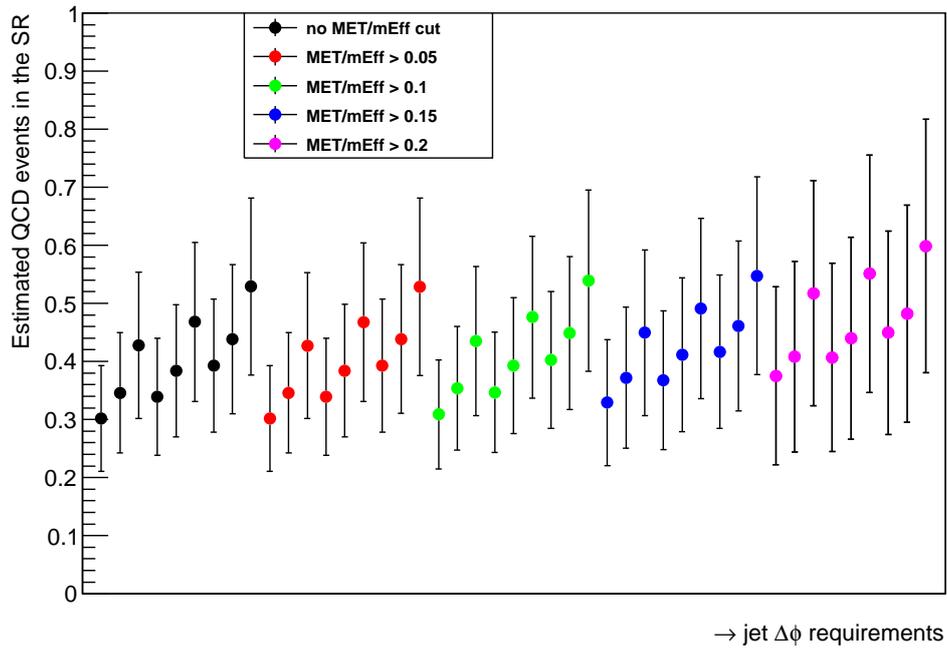


Figure 7.10: Estimated QCD background in the signal region for 45 different combinations of cuts defining the regions A, B, C and D. The different colored points correspond to different values of the lower cut on  $E_T^{\text{miss}}/m_{\text{eff}}$ . Within each group of colored points, the nine variations correspond to three variations on each of the leading and second-leading jet  $\Delta\phi$  cuts, varying the cut up to 0.35, using the nominal value of 0.3, and varying the cut down to 0.25. The nominal set of cuts corresponds to the fifth green point.

$$\hat{N}_{\text{SR}}^D = (0.4 \pm 0.7) \text{ events} . \quad (7.7)$$

This value is used for the results presented in chapter 9. It has a large uncertainty and can therefore only be regarded a “rough” estimate of the number of QCD-multijet background events. This is still sufficient, since this background is very small compared to the other Standard Model backgrounds in the analysis.



# Chapter 8 Systematic uncertainties

---

Up to this point no systematic uncertainties on the different physics objects have been considered. In this chapter the sources of systematic uncertainties are shortly explained and their impact is quantified for all backgrounds separately and the final estimates for the different backgrounds are given.

The systematic uncertainties on the following quantities were taken into account:

- Jet energy scale
- Jet energy resolution
- Tau energy scale
- Tau efficiency and misidentification probability
- $b$ -tagging efficiency
- Energy scale, resolution, efficiencies and misidentification probabilities of the light leptons

Most of these systematic uncertainties have been studied by the various combined performance groups within ATLAS which publish recommendations depending on the dataset and the physics objects.

## 8.1 Determination of the systematic uncertainties

For the QCD-multijet background, the systematic uncertainties are evaluated by applying the recommended variations to the different physics objects; this is done twice, independently for all the systematic uncertainties considered, shifting the value of the quantity at question by  $\pm 1\sigma$ . On these changed events the event selection for the signal region is applied and, to calculate the effect of each of the systematic uncertainties, the resulting number of events is

compared with the nominal event yield. To be less affected by possible statistical fluctuations the QCD-multijet sample, reweighted with misidentification probabilities is used.

For the  $W$ +jets and top contributions a different method is used. Instead of comparing the different yields in the signal region for the Monte Carlo samples, the effect of the systematic uncertainties on the transfer factors  $N_{\text{CR}_1}/N_{\text{SR}}$  and  $N_{\text{CR}_2}/N_{\text{SR}}$  (see chapter 6 for the definition of the two control regions), for events with identified tau leptons and events with misidentified tau candidates respectively, are studied.

The reason for this is, that some of the systematic uncertainties affect the event yield in the signal region in the same way as in the two control regions, hence these systematic uncertainties (partially) cancel when applying the transfer factor to the Monte Carlo expectation. With the transfer factors from the two control regions to the signal region the event yields can be compared to the ones where no systematic variation is applied.

The jet energy scale and tau energy scale are treated as fully correlated. This is a conservative assumption leading to larger uncertainties, since the variations of the two energy scales shift the transfer factors into the same direction.

Table 8.1 summarizes all uncertainties that go into the final estimates of the different backgrounds.

Both transfer factors depend on the fractions of  $W$ +jets and top in  $\text{CR}_1$  which is obtained by fitting the distribution of the BDT discriminant. To account for the changes due to the systematic variations in the obtained fractions, the fit to the BDT is repeated independently for each of the systematic variations. The changes to the scale factors for  $W$ +jets and top compared to the nominal result are shown in table 8.2.

In the following, a brief description of the different systematic uncertainties are given and it is explained how they are evaluated for the analysis.

### 8.1.1 Jet energy scale and resolution

The jet energy scale (JES) describes the relation between the true energy of a jet and the response of the calorimeter. The uncertainty on this relation has been studied by the ATLAS Jet/Etmiss Combined Performance Group and is described in detail in [42]. It is dependent on  $\eta$  and  $p_{\text{T}}$  of the jet. An additional correction is applied if a second jet is close-by [43]. The uncertainties are then applied to the event by rescaling the transverse momentum of all jets above  $p_{\text{T}} = 20 \text{ GeV}$  in a fully correlated way. The uncertainty on the jet energy scale ranges from below 2.5% for jets above  $p_{\text{T}} = 60 \text{ GeV}$  in the central calorimeter region ( $|\eta| < 0.8$ ) to 14% for jets between  $p_{\text{T}} = 20 \text{ GeV}$  and  $p_{\text{T}} = 30 \text{ GeV}$  in the forward region ( $3.2 < |\eta| < 4.5$ ). The jet energy scale uncertainty is also propagated to the missing transverse energy.

The corrections to the jet energies are applied before the jet-tau overlap removal. In this way the energy of all tau candidates is rescaled for the  $E_T^{\text{miss}}$  correction as well, which is appropriate since there is no dedicated term for tau leptons in the used definition of  $E_T^{\text{miss}}$  (see chapter 4) and tau candidates are treated as jets.

The uncertainty on the jet energy resolution arises from the finite precision of the detector simulation. The studies done by the Jet/Etmiss Combined Performance Group found the jet energy resolution to be in the order of 10% [44, 45]. The uncertainties provided by the Combined Performance Group are binned in  $p_T$  of the jets and they are applied by smearing the energies of all jets in the event randomly according to their uncertainty. The modified jet energies are propagated to  $E_T^{\text{miss}}$  in the same way as for the jet energy scale uncertainty.

### 8.1.2 Tau energy scale

For tau candidates an additional uncertainty on the energy scale to the one from the jet energy scale is applied due to the different calibration used. The uncertainty is dependent on  $\eta^\tau$ ,  $p_T^\tau$  and the number of tracks associated with the tau candidate and ranges from 3.5% to 9.5%. It is applied in the same way as the jet energy scale but only to tau candidates that pass the identification criteria.

Since the effect on the energy scale uncertainty for tau candidates on  $E_T^{\text{miss}}$  is already accounted for by the jet energy resolution no correction to  $E_T^{\text{miss}}$  for the tau energy scale uncertainty is done.

### 8.1.3 Tau misidentification probability rate and efficiency

The uncertainty on the tau misidentification probability and tau efficiency are still under study by the Tau Combined Performance Group.

For  $W$ +jets and top backgrounds, these uncertainties are already included in the uncertainty on the transfer factors from the two control regions to the signal region, thus no additional systematic uncertainty has to be applied.

For the QCD-multijet background the uncertainty on the tau misidentification probability was obtained with a dedicated dijet selection [32].

### 8.1.4 $b$ -tagging efficiency scale factor and uncertainty

The number of tagged  $b$ -jets in the event is used to separate  $W$ +jets from top events in section 6.1.2. The efficiency scale factor corrects for differences between data and simulation and is provided together with an associated uncertainty on the efficiency of the tagger used in the analysis by the Flavor Tagging Working Group at ATLAS. It is again dependent on

| Systematic [%]                      | top  | W+jets | QCD  |
|-------------------------------------|------|--------|------|
| Jet Energy Scale                    | 12.6 | 9.3    | 10.3 |
| Jet Energy Resolution               | 8.6  | 3.3    | 24.4 |
| $\tau$ Energy Scale                 | 4.3  | 3.9    | 13.2 |
| W/top fit stat.                     | 9.5  | 4.3    | –    |
| W/top fit syst.                     | 9.7  | 2.4    | –    |
| CR <sub>2</sub> statistical         | 7.8  | 7.8    | –    |
| CR <sub>2</sub> extrapolation to SR | 6.4  | 6.0    | –    |
| ABCD method                         | –    | –      | 35.2 |
| MC stat.                            | 20.9 | 19.5   | –    |
| Total                               | 31.5 | 25.1   | 46.0 |

Table 8.1: Systematic uncertainties in percent after the last cut step for W+jets, top and QCD. The total systematic line includes the JES-TES correlation and treats the true- and fake- $\tau$  contributions separately for the W+jets and top contributions and so is not simply the sum in quadrature of the individual uncertainties.

the jet kinematics and the flavor of the true quark, on which the efficiency of the  $b$ -tagging algorithm is strongly dependent on.

The scale factor ranges from 0.880 to 0.958 with an uncertainties of 7.7% to about 16%. The scale factor is applied as a weight to each event in the simulation.

### 8.1.5 Systematic uncertainties for light leptons

The impact of the systematic uncertainties of the light leptons on the number of selected events is also studied, but has been found to be negligible, their efficiencies and misidentification probabilities are a lot smaller compared to the tau leptons.

|                               | W+jets | top    |
|-------------------------------|--------|--------|
| <i>b</i> -tag efficiency up   | 0.024  | -0.097 |
| <i>b</i> -tag efficiency down | 0.024  | -0.097 |
| Jet energy resolution         | -0.012 | 0.032  |
| Jet energy scale up           | -0.134 | -0.090 |
| Jet Energy scale down         | 0.159  | 0.155  |
| Tau energy scale up           | 0.012  | 0.013  |
| Tau energy scale down         | 0.024  | 0.019  |
| Total                         | 0.212  | 0.229  |

Table 8.2: Relative deviations in the scale factors for the W+jets and top contributions, measured for different systematic variations.



# Chapter 9 Results

---

In this chapter, first an overview of the results of the signal selection using the background estimates from chapter 6 and 7 is given. The results are compared to data after the requirement on  $m_T(\tau, E_T^{\text{miss}})$  and in the signal region.

In section 6.1 two scale factors for the contributions from  $W$ +jets and top with identified tau leptons were derived. The scale factors with respect to the Monte Carlo prediction are:

$$f_W = 0.76 \pm 0.03 \quad (9.1)$$

$$f_{\text{top}} = 1.59 \pm 0.15. \quad (9.2)$$

Figure 9.1 shows the  $m_{\text{eff}}$  and the  $E_T^{\text{miss}}$  distributions in the true-tau control region  $\text{CR}_1$  before and after the scale factors are applied. A slight improvement between data and Monte Carlo is seen in both variables. To quantify the improvement a Kolmogorov-Smirnov test has been performed on both distributions. The K-S probability for the  $m_{\text{eff}}$  distribution increases from 0.97 to 0.99 and from 0.47 to 0.67 for the  $E_T^{\text{miss}}$  distribution.

Both variables were not used in the training of the BDT classifier, used to separate the  $W$ +jets and top backgrounds, to derive the two scale factors.

In a different control region,  $\text{CR}_2$ , a combined scale factors for  $W$ +jets, top and  $Z$ +jets events with misidentified tau candidates,  $f_{\text{fake}}$ , was derived (see section 6.2):

$$f_{\text{fake}} = 0.82 \pm 0.10. \quad (9.3)$$

Figure 9.2 shows the  $m_{\text{eff}}$  and  $E_T^{\text{miss}}$  distributions in  $\text{CR}_2$ . First without any scaling, then after the scale factors from  $\text{CR}_1$  are applied to the  $W$ +jets and top Monte Carlo samples and after the additional scale factor derived in  $\text{CR}_2$  is applied. Again, a slight improvement in the agreement between data and Monte Carlo is seen. As for the distributions in  $\text{CR}_1$  a Kolmogorov-Smirnov test was performed. For the  $m_{\text{eff}}$  distribution the K-S probability in-

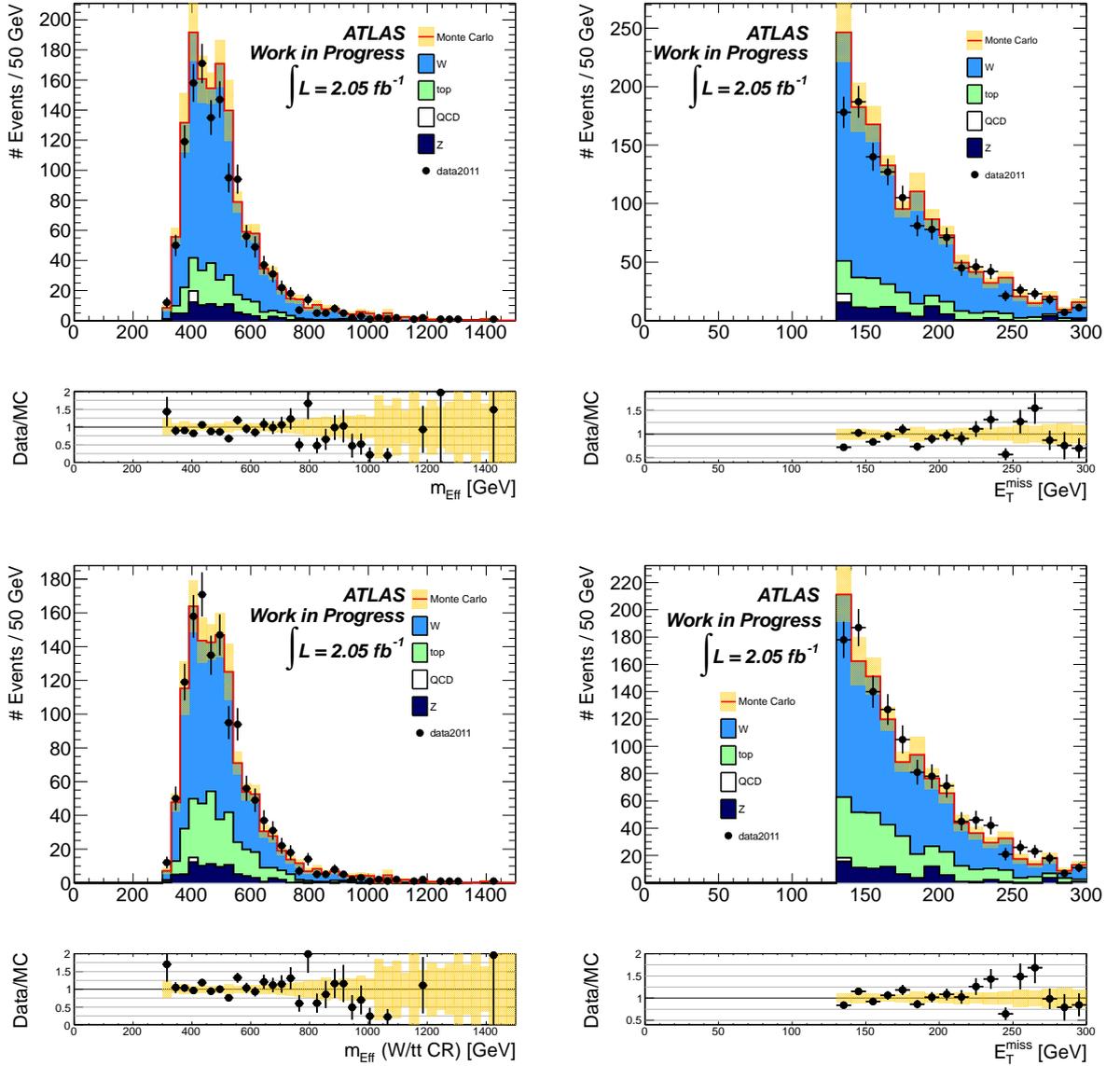


Figure 9.1:  $m_{\text{eff}}$  distribution (left column) and  $E_T^{\text{miss}}$  distribution (right column) in the true-tau control region  $\text{CR}_1$  before (top figure) and after (bottom figure) the  $W$ +jets and top scale factors are applied. The error band includes the statistical uncertainty on the sum of all backgrounds and the uncertainty on the scale factors (if applied).

| Requirement  | Data | top           | W+jets         | Z+jets         | QCD           | $\Sigma_{\text{SM}}$ |
|--|------|---------------|----------------|----------------|---------------|----------------------|
| $m_{\text{T}}(\tau, E_{\text{T}}^{\text{miss}}) > 110 \text{ GeV}$ | 96   | $45 \pm 13$   | $38 \pm 10$    | $34.3 \pm 8.4$ | $3.8 \pm 3.8$ | $121.1 \pm 32.8$     |
| $m_{\text{eff}} > 600 \text{ GeV}$                                 | 20   | $9.4 \pm 3.3$ | $10.1 \pm 3.3$ | $9.7 \pm 2.5$  | $0.4 \pm 0.7$ | $29.6 \pm 9.0$       |

Table 9.1: Numbers of events observed in data and estimated Standard Model backgrounds for an integrated luminosity of  $2.05 \text{ fb}^{-1}$ . The uncertainties include all statistical and systematic uncertainties.

creases from 0.64 (before scaling) to 0.68 (after scale factors from  $\text{CR}_1$  are applied) and 0.81 (after all scale factors are applied). The same has been done for the  $E_{\text{T}}^{\text{miss}}$  distribution. The K-S probabilities obtained are 0.87, 0.9 and 0.94 respectively.

In chapter 7, a data-driven estimation of the QCD-multijet background was presented. The final result for the number of QCD-multijet events in the signal region is:

$$\hat{N}_{\text{QCD}}^{\text{SR}} = (0.4 \pm 0.7) \text{ events} . \quad (9.4)$$

The number of observed data events and the estimated Standard Model backgrounds for an integrated luminosity of  $2.05 \text{ fb}^{-1}$  is summarized in table 9.1. The numbers include the statistical and systematic uncertainties discussed in chapter 8.

The W+jets and top estimates include the contributions of both correctly identified tau leptons and misidentified tau candidates. Both samples have a purity of correctly identified tau leptons of about 35% in both selection steps shown in table 9.1. The Z+jets and QCD-multijet backgrounds consist entirely of events with misidentified tau candidates.

The expected number of background events for Z+jets is based on simulation but is corrected with the scale factor for events with misidentified tau candidates derived in section 6.2, the W+jets and top backgrounds are based on data from the control regions described in chapter 6 and the number of QCD-multijet events is estimated from data in chapter 7.

Other backgrounds, mostly from diboson production have been found to amount to less than 0.2 events and are neglected for this analysis.

The expected signal yields for the  $\text{GMSB}_{\Lambda=30}^{\tan\beta=20}$  and  $\text{GMSB}_{\Lambda=40}^{\tan\beta=30}$  SUSY scenarios are 54.8 and 10.5 events after the requirement on  $m_{\text{T}}(\tau, E_{\text{T}}^{\text{miss}})$ , respectively, and 49.5 and 9.6 events after the requirement on  $m_{\text{eff}}$ . The observed event yields are consistent with the total expectation from Standard Model processes within slightly more than one standard deviation.

The distributions of the effective mass,  $E_{\text{T}}^{\text{miss}}$  and the  $p_{\text{T}}$  of the leading tau candidate and

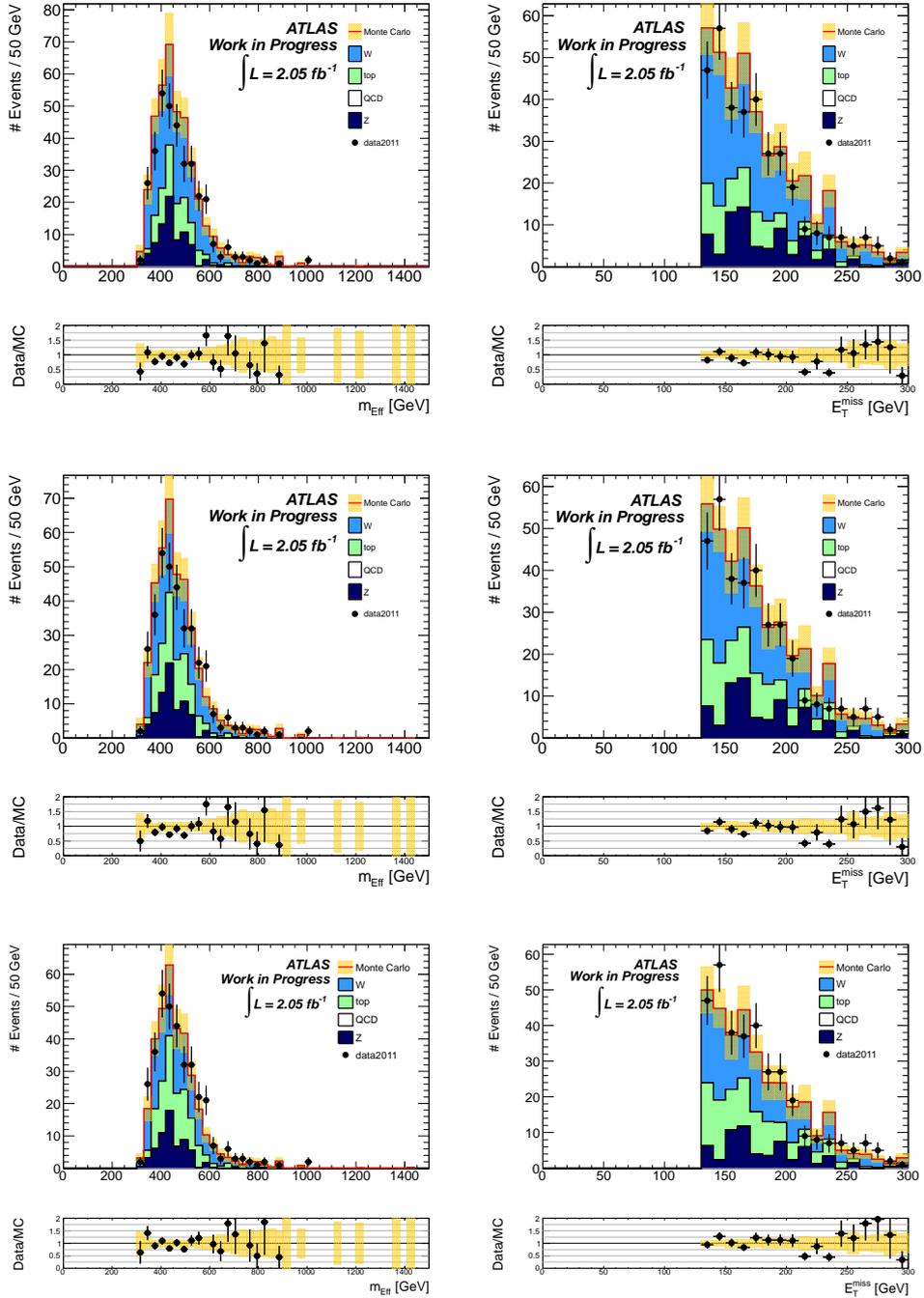


Figure 9.2:  $m_{\text{eff}}$  distribution (left column) and  $E_T^{\text{miss}}$  distribution (right column) in  $\text{CR}_2$  before any scaling is applied (top figure), after the scale factors from  $\text{CR}_1$  are applied to the  $W$ +jets and top samples (center figure) and after the combined scale factor ( $f_{\text{fake}}$ ) is applied to all samples (bottom figure).

The error band includes the statistical uncertainty on the sum of all backgrounds and the uncertainty on the scale factors (if applied).

the leading jet in the event are shown in figure 9.3 before the final requirement  $m_{\text{eff}} > 600 \text{ GeV}$ . For illustration purposes, the distributions of the two reference GMSB signal scenarios are also shown. A good agreement between data and the Monte Carlo expectation is seen for all variables. No excess of events that could be a sign of supersymmetry is observed thus exclusion limits are set.

## 9.1 Exclusion limits

The results from the signal selection and the background estimation are used to calculate an exclusion limit in the  $(\Lambda, \tan \beta)$  plane for the GMSB scenario (see section 2.2.2). It is based on the signal selection and background estimation presented in this thesis, but have been produced within the SUSY+ $\tau$  group at ATLAS.

To calculate 95% confidence level upper limits, a profile log likelihood ratio (LLR) test is used. The details of this method can be found in [46].

The general idea is to define a likelihood function  $L$  as:

$$L(n|s, b, \theta) = P_s(n|s, b) \times C_{\text{sys}}(\theta), \quad (9.5)$$

where  $n$  is number of observed data events,  $b$  the sum of all estimated backgrounds,  $s$  the number of expected events from the SUSY signal under study and  $\theta$  is a nuisance parameter describing the systematic uncertainties. The function  $P_s$  is a Poisson-probability distribution for the number of events in the signal region and  $C_{\text{sys}}$  represents the constraints on the systematic uncertainties.

From this likelihood the LLR is computed which can then be used to find the one-sided upper limit at 95% confidence level by inverting the signal model hypothesis test. The resulting exclusion limit obtained in the minimal GMSB model in the  $(\Lambda, \tan \beta)$  plane is shown in figure 9.4. The region at large  $\tan \beta$  and low  $\Lambda$  is theoretically excluded since it leads to tachyonic final states. The remaining parameter space is partitioned into regions with different types of NLSP particles. The limit set by the LEP experiments is also shown.

Values of  $\Lambda$  below 30 to 40 TeV are excluded, depending on  $\tan \beta$ .

In the context of the search for supersymmetry with two light leptons [47], an exclusion limit was also obtained for the GMSB scenario. When comparing the two limits, it can be seen that the two lepton analysis excludes larger values of  $\Lambda$  for smaller  $\tan \beta$  while the limit calculated in this analysis extend to larger  $\tan \beta$ , due to the enhanced coupling to the stau with increasing  $\tan \beta$ .

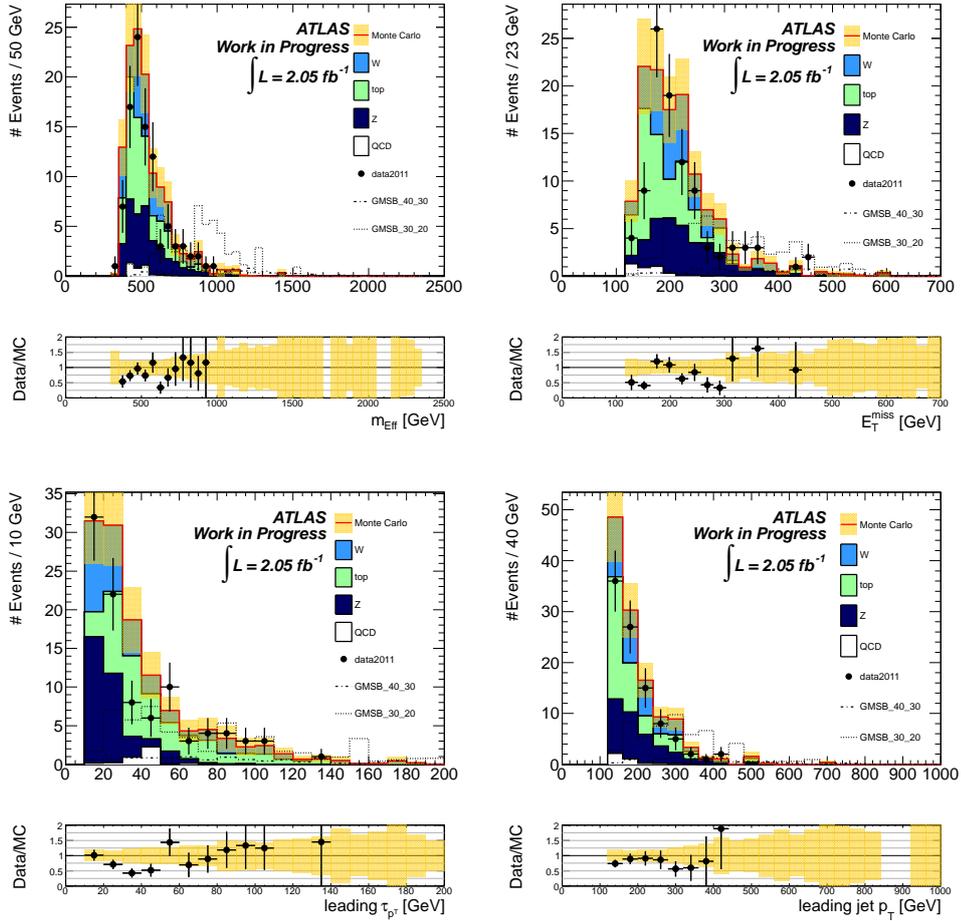


Figure 9.3: Distributions of the effective mass (upper left),  $E_T^{\text{miss}}$  (upper right), the  $p_T$  of the leading tau candidate (lower left) and the leading jet  $p_T$ . All distributions are shown before the final requirement on  $m_{\text{eff}}$ . The yellow band denotes the combination of systematic and statistical uncertainties for the sum of all Standard Model background. The error bars on the black data points denote only their Poissonian statistical uncertainties.

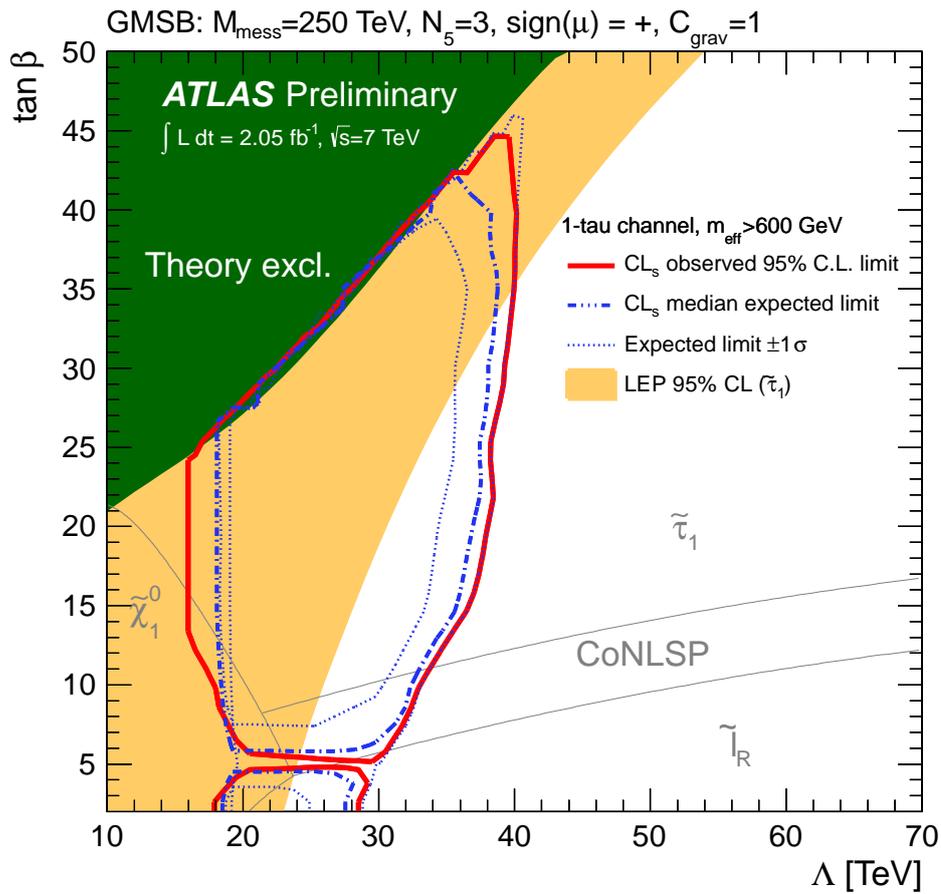


Figure 9.4: The observed and expected 95% confidence level exclusion limit in the  $(\Lambda, \tan \beta)$  plane for the minimal gauge mediated SUSY breaking model. The  $\pm 1\sigma$  expected exclusion curves are also shown.



# Chapter 10 Conclusion

---

In this thesis, techniques for estimating the  $W$ +jets, top and QCD-multijet background in the search for supersymmetric particles in final states with jets, missing transverse energy and tau leptons were presented. Data of  $pp$ -collisions recorded with the ATLAS detector at the LHC in 2011 at a center-of-mass energy of  $\sqrt{s} = 7$  TeV, corresponding to an integrated luminosity of  $\mathcal{L} = 2.05 \text{ fb}^{-1}$ , was used for the analysis.

Both processes with real tau leptons and real missing transverse energy and processes with misidentified tau candidates or missing transverse energy due to detector effects are backgrounds for this search.

The estimation of the  $W$ +jets and top backgrounds is divided into two parts: First the estimation of the number of  $W$ +jets and top events with real tau leptons was performed, using a  $W$ +jets and top enhanced control region. To separately estimate the two contributions, a Boosted Decision Tree that uses four characteristic variables, which separate between the contributions from  $W$ +jets and top, was trained to classify events in that control region. By comparing the expectations from simulation to data, scale factors for the  $W$ +jets and top backgrounds were derived.

The combined number of  $W$ +jets and top background events with misidentified tau candidates was estimated in a second control region with enhanced contributions from events with misidentified tau candidates. By subtracting off the contributions from  $W$ +jets and top events with real tau leptons from data and comparing the remaining number of events with the expectation from simulation, a combined scale factors for events with misidentified tau candidates was obtained.

The QCD-multijet background was estimated using a matrix-method. Instead of asking for a tau candidate fulfilling all identification criteria, one tau candidate with only loose preselection criteria was chosen randomly. The obtained sample was then split up in four regions. The method only provides a rough estimate of the number of QCD-multijet events in the signal region, which is sufficient, since it is seen that this background is very small compared to the other backgrounds in the analysis.

20 data events have been observed in the signal region with an estimated sum of Standard Model backgrounds of  $(29.6 \pm 9.0)$  events. A good agreement between the data and the Standard Model expectation is seen, thus the event selection and the background estimates presented in this thesis were used to set exclusion limits in the  $(\Lambda, \tan \beta)$ -plane in the GMSB scenario. These exclusion limits are more stringent than limits from previous experiments.

In the future the systematic uncertainties on the various physics objects are expected to decrease significantly. In addition to that, with a bigger dataset the statistical precision in all control regions can be drastically increased. Both leads to more precise estimates on all contributing backgrounds and subsequently better sensitivity for supersymmetric scenarios not excluded in this analysis.

# Appendix A Used Monte Carlo samples

---

| Sample ID | Name           | Generator    | $\epsilon_f \cdot \sigma_{\text{LO}}$ [pb] | k-factor | $\epsilon_f \cdot \sigma_{\text{NLO}}$ [pb] |
|-----------|----------------|--------------|--|----------|---|
| 105200    | ttbar T1       | MC@NLO/Jimmy | 164.57                                     | 0.543    | 89.36                                       |
| 105204    | ttbarhad       | MC@NLO/Jimmy | 164.57                                     | 0.457    | 75.21                                       |
| 108340    | st tchan enu   | MC@NLO/Jimmy | 7.12                                       | 1.0      | 7.12  |
| 108341    | st tchan munu  | MC@NLO/Jimmy | 7.12                                       | 1.0      | 7.12  |
| 108342    | st tchan taunu | MC@NLO/Jimmy | 7.10                                       | 1.0      | 7.10  |
| 108343    | st schan enu   | MC@NLO/Jimmy | 0.47                                       | 1.0      | 0.47  |
| 108344    | st schan munu  | MC@NLO/Jimmy | 0.47                                       | 1.0      | 0.47  |
| 108345    | st schan taunu | MC@NLO/Jimmy | 0.47                                       | 1.0      | 0.47  |
| 108346    | st Wt          | MC@NLO/Jimmy | 14.59                                      | 1.0      | 14.59                                       |

Table A.1: Used  $t\bar{t}$  and single  $t$  Monte Carlo samples with their corresponding sample IDs, event generators, and filter efficiency times cross-sections.

| Sample ID | Name     | Generator    | $\epsilon_f \cdot \sigma_{\text{LO}}$ [pb] | k-factor | $\epsilon_f \cdot \sigma_{\text{NNLO}}$ [pb] |
|-----------|----------|--------------|--|----------|--|
| 107680    | Wenu0p   | Aplgen/Jimmy | 6921.6                                     | 1.20     | 8305.92                                      |
| 107681    | Wenu1p   | Aplgen/Jimmy | 1304.3                                     | 1.20     | 1565.16                                      |
| 107682    | Wenu2p   | Aplgen/Jimmy | 378.3                                      | 1.20     | 453.95                                       |
| 107683    | Wenu3p   | Aplgen/Jimmy | 101.4                                      | 1.20     | 121.72                                       |
| 107684    | Wenu4p   | Aplgen/Jimmy | 25.9                                       | 1.20     | 31.04  |
| 107685    | Wenu5p   | Aplgen/Jimmy | 7.0  | 1.20     | 8.40   |
| 107690    | Wmunu0p  | Aplgen/Jimmy | 6919.6                                     | 1.20     | 8303.52                                      |
| 107691    | Wmunu1p  | Aplgen/Jimmy | 1304.2                                     | 1.20     | 1565.04                                      |
| 107692    | Wmunu2p  | Aplgen/Jimmy | 377.8                                      | 1.20     | 453.39                                       |
| 107693    | Wmunu3p  | Aplgen/Jimmy | 101.9                                      | 1.20     | 122.26                                       |
| 107694    | Wmunu4p  | Aplgen/Jimmy | 25.8                                       | 1.20     | 30.90  |
| 107695    | Wmunu5p  | Aplgen/Jimmy | 6.9  | 1.20     | 8.30   |
| 107700    | Wtaunu0p | Aplgen/Jimmy | 6918.6                                     | 1.20     | 8302.32                                      |
| 107701    | Wtaunu1p | Aplgen/Jimmy | 1303.2                                     | 1.20     | 1563.84                                      |
| 107702    | Wtaunu2p | Aplgen/Jimmy | 378.2                                      | 1.20     | 453.82                                       |
| 107703    | Wtaunu3p | Aplgen/Jimmy | 101.5                                      | 1.20     | 121.81                                       |
| 107704    | Wtaunu4p | Aplgen/Jimmy | 25.6                                       | 1.20     | 30.77  |
| 107705    | Wtaunu5p | Aplgen/Jimmy | 7.0  | 1.20     | 8.45   |

Table A.2: Used  $W$ +jets Monte Carlo samples with their corresponding sample IDs, event generators, filter efficiency times cross-sections, and k-factors.

| Sample ID | Name      | Generator    | $\epsilon_f \cdot \sigma_{\text{LO}}$ | k-factor | $\epsilon_f \cdot \sigma_{\text{NNLO}}$ [pb] |
|-----------|-----------|--------------|---------------------------------------|----------|--|
| 107650    | Zee0p     | Aplgen/Jimmy | 668.3                                 | 1.25     | 835.40                                       |
| 107651    | Zee1p     | Alpgen/Jimmy | 134.4                                 | 1.25     | 167.95                                       |
| 107652    | Zee2p     | Alpgen/Jimmy | 40.54                                 | 1.25     | 50.68  |
| 107653    | Zee3p     | Alpgen/Jimmy | 11.16                                 | 1.25     | 13.95  |
| 107654    | Zee4p     | Alpgen/Jimmy | 2.88                                  | 1.25     | 3.60   |
| 107655    | Zee5p     | Alpgen/Jimmy | 0.83                                  | 1.25     | 1.04   |
| 107660    | Zmumu0p   | Alpgen/Jimmy | 668.7                                 | 1.25     | 835.85                                       |
| 107661    | Zmumu1p   | Alpgen/Jimmy | 134.1                                 | 1.25     | 167.68                                       |
| 107662    | Zmumu2p   | Alpgen/Jimmy | 40.33                                 | 1.25     | 50.41  |
| 107663    | Zmumu3p   | Alpgen/Jimmy | 11.19                                 | 1.25     | 13.99  |
| 107664    | Zmumu4p   | Alpgen/Jimmy | 2.75                                  | 1.25     | 3.44   |
| 107665    | Zmumu5p   | Alpgen/Jimmy | 0.77                                  | 1.25     | 0.96   |
| 107670    | Ztautau0p | Alpgen/Jimmy | 668.4                                 | 1.25     | 835.50                                       |
| 107671    | Ztautau1p | Alpgen/Jimmy | 134.8                                 | 1.25     | 168.51                                       |
| 107672    | Ztautau2p | Alpgen/Jimmy | 40.36                                 | 1.25     | 50.45  |
| 107673    | Ztautau3p | Alpgen/Jimmy | 11.25                                 | 1.25     | 14.06  |
| 107674    | Ztautau4p | Alpgen/Jimmy | 2.79                                  | 1.25     | 3.49   |
| 107675    | Ztautau5p | Alpgen/Jimmy | 0.77                                  | 1.25     | 0.96   |
| 107710    | ZnunuNp0  | Alpgen/Jimmy | 26.71                                 | 1.282    | 34.22  |
| 107711    | ZnunuNp1  | Alpgen/Jimmy | 451.4                                 | 1.282    | 578.54                                       |
| 107712    | ZnunuNp2  | Alpgen/Jimmy | 197.6                                 | 1.282    | 253.29                                       |
| 107713    | ZnunuNp3  | Alpgen/Jimmy | 59.89                                 | 1.282    | 76.75  |
| 107714    | ZnunuNp4  | Alpgen/Jimmy | 15.61                                 | 1.282    | 20.01  |
| 107715    | ZnunuNp5  | Alpgen/Jimmy | 4.17                                  | 1.282    | 5.34   |

Table A.3: Used Z+jets Monte Carlo samples with their corresponding sample IDs, event generators, filter efficiency times cross-sections, and k-factors.

| Sample ID | Name | Generator | $\epsilon_f \cdot \sigma_{\text{LO}}$ [pb] |
|-----------|------|-----------|--|
| 105009    | J0   | Pythia    | 9860800000                                 |
| 105010    | J1   | Pythia    | 678180000                                  |
| 105011    | J2   | Pythia    | 40982000                                   |
| 105012    | J3   | Pythia    | 2192900                                    |
| 105013    | J4   | Pythia    | 87701                                      |
| 105014    | J5   | Pythia    | 2350.1                                     |
| 105015    | J6   | Pythia    | 33.61                                      |
| 105016    | J7   | Pythia    | 0.13744                                    |
| 105017    | J8   | Pythia    | 0.0000062                                  |

Table A.4: Used dijet Monte Carlo samples with their corresponding sample IDs, event generators, and filter efficiency times cross-sections.

# Bibliography

---

- [1] D. Griffiths. *Introduction to Elementary Particles*. 2nd and revised. Wiley-VCH, 2008.
- [2] B. Povh et al. *Teilchen und Kerne: Eine Einführung in die physikalischen Konzepte*. Springer, 2009.
- [3] T. Fließbach. *Elektrodynamik: Lehrbuch Zur Theoretischen Physik II*. Lehrbuch zur theoretischen Physik / Torsten Fließbach. Spektrum Akademischer Verlag, 2008.
- [4] K. Nakamura et al (Particle Data Group). 'Review of Particle Physics, 2010-2011. Review of Particle Properties'. In: *J. Phys. G* 37.7A (2010), p. 075021.
- [5] *Combination of the Searches for the Higgs Boson in  $\sim 1\text{fb}^{-1}$  of Data Taken with the ATLAS Detector at 7 TeV Center-of-Mass Energy*. Tech. rep. ATLAS-CONF-2011-112. Geneva: CERN, Aug. 2011.
- [6] M. E. Peskin. *Beyond the standard model*. 1997. arXiv:hep-ph/9705479 [hep-ph].
- [7] S. P. Martin. 'A Supersymmetry Primer'. In: *Phys. Lett.* B393 (Sept. 2001). eprint: hep-ph/9709356v6.
- [8] H. Baer and X. Tata. *Weak Scale Supersymmetry*. Cambridge University Press, 2006.
- [9] L. Evans and P. Bryant. 'LHC Machine'. In: *Journal of Instrumentation* 3.08 (2008), S08001. URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08001>.
- [10] The ATLAS Collaboration. 'The ATLAS Experiment at the CERN Large Hadron Collider'. In: *Journal of Instrumentation* 3.08 (2008), S08003. URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08003>.
- [11] The CMS Collaboration. 'The CMS experiment at the CERN LHC'. In: *Journal of Instrumentation* 3.08 (2008), S08004. URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08004>.
- [12] The LHCb Collaboration. 'The LHCb Detector at the LHC'. In: *Journal of Instrumentation* 3.08 (2008), S08005. URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08005>.
- [13] The ALICE Collaboration. 'The ALICE experiment at the CERN LHC'. In: *Journal of Instrumentation* 3.08 (2008), S08002.

- [14] The LHCf Collaboration. 'The LHCf detector at the CERN Large Hadron Collider'. In: *Journal of Instrumentation* 3.08 (2008), S08006. URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08006>.
- [15] The TOTEM Collaboration. 'The TOTEM Experiment at the CERN Large Hadron Collider'. In: *Journal of Instrumentation* 3.08 (2008), S08007. URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08007>.
- [16] AC Team. 'The four main LHC experiments'. June 1999.
- [17] CERN AC. 'Layout of ATLAS. Dessin representant le detecteur ATLAS'. Mar. 1998.
- [18] G. Aad et al. *Expected performance of the ATLAS experiment: detector, trigger and physics*. Geneva: CERN, 2009.
- [19] J. Pequeno. 'Computer generated image of the ATLAS inner detector'. Mar. 2008.
- [20] G. Aad et al. 'ATLAS pixel detector electronics and sensors'. In: *Journal of Instrumentation* 3.07 (2008), P07007. URL: <http://stacks.iop.org/1748-0221/3/i=07/a=P07007>.
- [21] The ATLAS TRT collaboration. 'The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: design and performance'. In: *Journal of Instrumentation* 3.02 (2008), P02013. URL: <http://stacks.iop.org/1748-0221/3/i=02/a=P02013>.
- [22] C. Grupen. *Teilchendetektoren*. BI-Wissenschaftsverlag, 1993.
- [23] J. Pequeno. 'Event Cross Section in a computer generated image of the ATLAS detector.' Mar. 2008.
- [24] J. Pequeno. 'Computer Generated image of the ATLAS calorimeter'. Mar. 2008.
- [25] A. Baden. 'Jets and kinematics in hadronic collisions'. In: *Int.J.Mod.Phys.* A13 (1998), pp. 1817–1845.
- [26] G. Piacquadio and C. Weiser. 'A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS'. In: *J. Phys. Conf. Ser.* 119 (2008), p. 032032.
- [27] 'Twiki page on b-tagging with JetFitter'. URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/BTaggingJetFitter>.
- [28] *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*. Tech. rep. ATLAS-CONF-2011-102. Geneva: CERN, July 2011.
- [29] E. Barberio et al. *Hadronic Tau Identification Performance and Efficiency Measurement*. Tech. rep. ATLAS-COM-CONF-2011-179. Geneva: CERN, Oct. 2011.
- [30] G. Aad et al. 'Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS'. In: (2011). arXiv:1108.5602 [hep-ex].
- [31] D. R. Tovey. 'Measuring the SUSY Mass Scale at the LHC'. In: *PHYS.LETT.B* 498 (2001).

- [32] M. Mazur et al. *Search for supersymmetry with jets, missing transverse momentum, and taus at  $\sqrt{s} = 7$  TeV: Support note for 2011 analysis*. Tech. rep. ATL-COM-PHYS-2011-964. Geneva: CERN, July 2011.
- [33] The Luminosity Group. *Luminosity Determination in pp Collisions at  $\sqrt{s} = 7$  TeV using the ATLAS Detector in 2011*. Tech. rep. ATLAS-COM-CONF-2011-130. Geneva: CERN, July 2011.
- [34] S. Asai et al. *Further search for squarks and gluinos using final states with jets and missing transverse momentum with the ATLAS experiment in  $\sqrt{s} = 7$  TeV proton-proton collisions: supporting documentation*. Tech. rep. ATL-COM-PHYS-2011-784. Supporting documentation for 2011 SUSY 0-lepton EPS paper. Geneva: CERN, June 2011.
- [35] 'Homepage of the ATLAS run query'. URL: <http://atlas-runquery.cern.ch>.
- [36] 'Twiki page on the definition of bad jets'. URL: [https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HowToCleanJets#Bad\\_jets\\_rel16\\_data](https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HowToCleanJets#Bad_jets_rel16_data).
- [37] *Measurement of the Mis-identification Probability of Tau Leptons from Hadronic Jets and from Electrons*. Tech. rep. ATLAS-CONF-2011-113. Geneva: CERN, July 2011.
- [38] J. Therhaag. 'Advanced Boosting in TMVA4, Presented at the DPG Frühjahrstagung 2010, Bonn'. 2010.
- [39] 'Homepage of the TMVA project'. URL: <http://www.tmva.sourceforge.net>.
- [40] R. Barlow and C. Beeston. 'Fitting using finite Monte Carlo samples'. In: *Comp. Phys. Comm.* 77 (1993), p. 219.
- [41] 'Homepage of the ROOT analysis framework'. URL: <http://root.cern.ch>.
- [42] *Jet energy scale and its systematic uncertainty in proton-proton collisions at  $\sqrt{s}=7$  TeV in ATLAS 2010 data*. Tech. rep. ATLAS-CONF-2011-032. Geneva: CERN, Mar. 2011.
- [43] *Close-by Jet Effects on Jet Energy Scale Calibration in pp Collisions at  $\sqrt{s}=7$  TeV with the ATLAS Detector*. Tech. rep. ATLAS-CONF-2011-062. Geneva: CERN, Apr. 2011.
- [44] *Jet energy resolution and selection efficiency relative to track jets from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy  $\sqrt{s} = 7$  TeV*. Tech. rep. ATLAS-CONF-2010-054. Geneva: CERN, July 2010.
- [45] G. Romeo et al. *Jet Energy Resolution from In-situ Techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy  $\sqrt{s} = 7$  TeV*. Tech. rep. ATL-COM-PHYS-2011-240. Geneva: CERN, Mar. 2011.
- [46] R. Bruneliere et al. *Setting exclusion limits in ATLAS supersymmetry searches with a likelihood ratio based method*. Tech. rep. ATL-PHYS-INT-2011-032. Geneva: CERN, Apr. 2011.

- [47] A. Alonso et al. *Search for Supersymmetry with two leptons and missing transverse momentum at  $\sqrt{s} = 7$  TeV: Interpretation in the GMSB scenario*. Tech. rep. ATL-COM-PHYS-2011-1381. Geneva: CERN, Oct. 2011.

# List of Figures

---

|      |   |    |
|------|---|----|
| 2.1  | The elementary vertex of QED. . . . .   | 7  |
| 2.2  | Feynman diagrams for electron-muon scattering. . . . .  | 8  |
| 2.3  | Possible vertices of QCD. . . . .   | 8  |
| 2.4  | Vertices of the weak interaction. . . . .   | 9  |
| 2.5  | Graph of $U(\phi)$ . . . . .  | 10 |
| 2.6  | Feynman diagram for the decay of a tau lepton into light leptons or quarks. . . . .   | 11 |
| 2.7  | Contribution to the Higgs mass from a loop diagram involving the top quark. . . . .   | 13 |
| 2.8  | Running of the coupling constants in the SM. . . . .  | 14 |
| 2.9  | Energy density of the universe . . . . .  | 15 |
| 2.10 | Possible Feynman Diagram of a proton-decay if both lepton and baryon number-violating terms are allowed. . . . .                                  | 18 |
| 2.11 | Contribution to the Higgs mass from a loop involving the superpartner of the top quark, the <i>stop</i> $\tilde{t}$ . . . . .                     | 19 |
| 2.12 | Running of the coupling constants in SUSY. . . . .  | 20 |
| 2.13 | schematic structure for SUSY breaking. . . . .  | 21 |
| 2.14 | Example for a decay of a $\tilde{q}$ involving two tau leptons. . . . .   | 23 |
| 2.15 | Feynman Diagram of the production of two squarks in a $pp$ -collision. . . . .  | 24 |
| 3.1  | Schematic view of the Large Hadron Collider with the four large experiments. . . . .  | 26 |
| 3.2  | Schematic view of the full ATLAS detector. . . . .  | 27 |
| 3.3  | The ATLAS inner detector components. . . . .  | 28 |
| 3.4  | Image of the ATLAS calorimeter system. . . . .  | 29 |
| 3.5  | Schematic cross-section of the ATLAS detector showing the interaction of different types of particles in the various detector components. . . . . | 31 |
| 4.1  | Sketch of the definition of jets. . . . .   | 34 |
| 5.1  | Trigger efficiency for the jet+ $E_T^{\text{miss}}$ -trigger used in the analysis. . . . .  | 41 |
| 5.2  | Number of identified electrons and muons. . . . .   | 43 |

|      |  |    |
|------|--|----|
| 5.3  | Multiplicity of jets with $p_T > 30 \text{ GeV}$ after the trigger requirements and the rejection of events with identified light leptons. . . . .   | 43 |
| 5.4  | Number of tau candidates after all other selection criteria are applied. . . . .   | 44 |
| 5.5  | Distributions of QCD-rejection variables . . . . .   | 45 |
| 5.6  | Distribution of $m_T(\tau, E_T^{\text{miss}})$ after QCD suppression . . . . .   | 46 |
| 5.7  | Distribution of the effective mass $m_{\text{eff}}$ , shown after all other selection criteria are applied. . . . .  | 47 |
| 5.8  | Tau misidentification probabilities for the used $\text{BDT}_{\text{tau}}$ working point. . . . .  | 49 |
| 6.1  | $m_T(\tau, E_T^{\text{miss}})$ split up in events with identified and misidentified tau candidates. . . . .  | 52 |
| 6.2  | Definition of the two control regions ( $\text{CR}_1$ & $\text{CR}_2$ ) and the signal region (SR) in the $((m_T(\tau, E_T^{\text{miss}}) - m_{\text{eff}})$ -plane. . . . .                                 | 53 |
| 6.3  | Number of events in the $(E_T^{\text{miss}}/m_{\text{eff}}, \Delta\phi(E_T^{\text{miss}}, \text{jet}_{1,2}))$ -plane for the $W$ +jets, top and QCD-multijet samples from simulations and from data. . . . . | 56 |
| 6.4  | $\Delta\phi(E_T^{\text{miss}}, \text{jet}_{1,2})$ distribution . . . . .   | 57 |
| 6.5  | Kinematic distributions in the true-tau control region $\text{CR}_1$ . The errors shown are statistical only. . . . .  | 59 |
| 6.6  | Feynman diagram for the decay of the top quark. . . . .  | 60 |
| 6.7  | Example feynman diagrams for the production of a $W$ boson in association with 1 and 2 jets. . . . .   | 60 |
| 6.8  | Input variables to the classification of $W$ +jets and top events shown in $\text{CR}_1$ . . . . .   | 61 |
| 6.9  | Schematic view of a Decision tree and corresponding regions of the input parameter space . . . . .   | 62 |
| 6.10 | BDT discriminant computed from the input variables in $\text{CR}_1$ . . . . .  | 63 |
| 6.11 | Output of the Barlow-Beeston fit to the BDT distribution in data . . . . .   | 66 |
| 6.12 | Output of the Barlow-Beeston fit to the BDT distribution in data, trained without the information of the number of $b$ -jets . . . . .   | 67 |
| 6.13 | Evolution of the uncertainty on the fit parameters for different scalings. . . . .   | 69 |
| 6.14 | Result of 10000 fits of the input templates to different sets of toy-data. . . . .   | 70 |
| 6.15 | BDT input variables after scale factors applied . . . . .  | 71 |
| 6.16 | Comparison of the sources of misidentified tau candidates between $\text{CR}_2$ and the SR . . . . .   | 73 |
| 6.17 | Comparison of the $p_T$ of the misidentified tau candidates between $\text{CR}_2$ and the SR. . . . .  | 74 |
| 7.1  | Schematic view of the control regions (CR) <b>A–C</b> and the signal region (SR) <b>D</b> in the $\theta_1$ - $\theta_2$ paramter space. . . . .   | 76 |
| 7.2  | Distribution of $E_T^{\text{miss}}/m_{\text{eff}}$ in control region C. . . . .  | 80 |

|      |  |     |
|------|--|-----|
| 7.3  | $m_{\text{eff}}$ distribution in the for different control regions of the QCD-multijet background estimation method . . . . .  | 82  |
| 7.4  | $E_{\text{T}}^{\text{miss}}/m_{\text{eff}}$ distribution in control region A. This distribution is used to calculate the extrapolation factor from the intermediate $E_{\text{T}}^{\text{miss}}/m_{\text{eff}}$ region to the signal region. . . . . | 83  |
| 7.5  | Comparison of the $m_{\text{eff}}$ distributions for QCD Monte Carlo and the QCD estimate from data. . . . .   | 84  |
| 7.6  | Distributions of the tau identification $\text{BDT}_{\text{tau}}$ in slices of $\Delta\phi$ . . . . .  | 85  |
| 7.7  | Comparison of the shapes between the $m_{\text{eff}}$ -distributions in the combined regions A+B and C+D and regions A+C and B+D . . . . .   | 86  |
| 7.8  | $E_{\text{T}}^{\text{miss}}/m_{\text{eff}}$ distribution in control region B. . . . .  | 87  |
| 7.9  | Results of 10000 toy MC experiments validating the matrix method for QCD-multijet background estimation . . . . .  | 89  |
| 7.10 | Estimated QCD background in the signal region for 45 different combinations of cuts defining the regions A, B, C and D. . . . .  | 90  |
| 9.1  | Distributions in $\text{CR}_1$ with and without scaling applied. . . . .   | 100 |
| 9.2  | Distributions in $\text{CR}_2$ with and without scaling applied. . . . .   | 102 |
| 9.3  | Distributions before requirement on $m_{\text{eff}}$ is applied. . . . .   | 104 |
| 9.4  | The observed and expected 95% confidence level exclusion limit in the $(\Lambda, \tan\beta)$ plane for the minimal gauge mediated SUSY breaking model. . . . .   | 105 |



# List of Tables

---

|     |   |    |
|-----|---|----|
| 2.1 | Overview of the particle spectrum of the Standard Model. . . . .  | 4  |
| 2.2 | Interactions in the Standard Model . . . . .  | 5  |
| 2.3 | Most prominent hadronic decay-channels of the tau lepton and their branching fractions. . . . .   | 12 |
| 2.4 | Chiral supermultiplets in the MSSM. . . . .   | 17 |
| 2.5 | Gauge supermultiplets in the MSSM. . . . .  | 17 |
| 5.1 | Used Monte Carlo background samples. . . . .  | 40 |
| 5.2 | Used Monte Carlo signal samples. . . . .  | 40 |
| 5.3 | Expected numbers of events from Monte Carlo simulation, normalized to $2.05 \text{ fb}^{-1}$ . . . . .  | 48 |
| 5.4 | Comparison of the QCD multijet event yields for the Monte Carlos samples with and without tau misidentification probability reweighting. . . . .    | 50 |
| 6.1 | Number of events for data and background simulation in the low- $m_T$ ( $\tau, E_T^{\text{miss}}$ ) control region $\text{CR}_1$ . . . . .          | 52 |
| 6.2 | Numbers of observed and expected events in $\text{QCD}_1$ . . . . .   | 54 |
| 6.3 | Numbers of observed and expected events in the zero-tau region $\text{QCD}_2$ . . . . .   | 54 |
| 6.4 | Transfer factors derived in the three subsamples defined in section 6.1.1. . . . .  | 57 |
| 6.5 | Results of two fits in $\text{CR}_1$ compared to MC predictions. . . . .  | 67 |
| 6.6 | Numbers of observed and expected events in $\text{CR}_2$ . . . . .  | 72 |
| 7.1 | Definition of signal and control regions for the QCD-multijet background estimation method. . . . .   | 78 |
| 7.2 | Numbers of data and simulated events in the four regions of the QCD background estimation method. . . . .   | 78 |
| 7.3 | Numbers of data and simulated events in the four regions of the QCD background estimation method with the additional modifications applied. . . . . | 81 |
| 8.1 | Systematic uncertainties for the different backgrounds. . . . .   | 96 |

|     |   |     |
|-----|---|-----|
| 8.2 | Relative deviations in the scale factors for the $W$ +jets and top contributions, measured for different systematic variations. . . . .                       | 97  |
| 9.1 | Numbers of events observed in data and estimated Standard Model backgrounds. . . . .  | 101 |
| A.1 | Used $t\bar{t}$ and single $t$ Monte Carlo samples with their corresponding sample IDs, event generators, and filter efficiency times cross-sections. . . . . | 109 |
| A.2 | Used $W$ +jets Monte Carlo samples with their corresponding sample IDs, event generators, filter efficiency times cross-sections, and k-factors. . . . .      | 110 |
| A.3 | Used $Z$ +jets Monte Carlo samples with their corresponding sample IDs, event generators, filter efficiency times cross-sections, and k-factors. . . . .      | 111 |
| A.4 | Used dijet Monte Carlo samples with their corresponding sample IDs, event generators, and filter efficiency times cross-sections. . . . .                     | 112 |

# Acknowledgments

---

I would like to thank everybody who directly or indirectly helped to complete this thesis. Michel and Michael for their endless support and their willingness to convince me more than once that I've overlooked something, Karl and Jochen for giving me the opportunity to graduate in two groups without ever feeling in-between and all the other members of the combined third floor/Westbau high energy working group for the nice atmosphere, foosball-matches, dinners, cakes, barbecues and postponed group excursions.

Zuletzt möchte ich mich bei all denen bedanken, die mich in den letzten sechs Jahren durch das Studium begleitet haben. Allen voran Anna, Adrian und Flo ohne die ich sicherlich niemals rausgefunden hätte wie sich so ein Physikstudium im dritten Semester eigentlich anfühlt und jetzt wahrscheinlich "irgendwas mit Medien" machen würde. Ein großer Dank gebührt natürlich auch meiner Freundin Steffi, meinen Eltern und meinen Geschwistern die immer zu mir gestanden haben, auch wenn es mal nicht so lief wie geplant.



# Erklärung zur Selbständigkeit

---

Hiermit erkläre ich, dass ich die vorgelegte Arbeit selbständig verfasst habe, sowie nur die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich und sinngemäß übernommene Zitate habe ich als solche gekennzeichnet.

Freiburg, den 1. Dezember 2011

Felix Bühner